Office of Naval Research
Contract N00014-67-A-0298-0006 NR-372-012

11

ESTIMATION FOR ROTATIONAL PROCESSES WITH ONE DEGREE OF FREEDOM

VE RI TAS

By

James Ting-Ho Lo and Alan S. Willsky

July 1972

Technical Report No. 635

Division of Engineering and Applied Physics
Harvard University · Cambridge, Massachusetts

104

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Division of Engineering and Applied Physics Harvard University Cambridge, Massachusetts	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

ESTIMATION FOR ROTATIONAL PROCESSES WITH ONE DEGREE OF FREEDOM

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Interim technical report

5. AUTHOR(S) (First name, middle initial, last name)

James T. Lo and Alan S. Willsky

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July 1972	104	47

8a. CONTRACT OR GRANT NO. N00014-67-A-0298-0006 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 635 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted by the U. S. Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research

13. ABSTRACT

A class of bilinear estimation problems involving single-degree-of-freedom rotation is formulated and resolved. Both continuous and discrete time estimation problems are considered. Error criteria, probability distributions, and optimal estimates on the circle are studied. An effective synthesis procedure for continuous time estimation is provided, and a generalization to estimation on arbitrary abelian Lie groups is included. An intrinsic difference between the discrete and continuous problems is discussed, and the complexity of the equations in the discrete time case is analyzed in this setting. Applications of these results to a number of practical problems including FM demodulation and frequency stability are examined.

ia

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0102-014-6700

Unclassified

Security Classification

D-23822

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Rotational Processes						
Single Degree of Freedom						
Bilinear Equations						
Error Criteria						
Optimal Estimation						
Folded Normal Distribution						
Conditional Probability						
Fourier Analysis						
Frequency Demodulation						
Frequency Stability						
Abelian Lie Groups						

ib

Office of Naval Research

Contract N00014-67-A-0298-0006

NR-372-012

ESTIMATION FOR ROTATIONAL PROCESSES WITH ONE DEGREE OF FREEDOM

By

James Ting-Ho Lo and Alan S. Willsky

Technical Report No. 635

July 1972

The research reported in this document was made possible through support extended the Division of Engineering and Applied Physics, Harvard University by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research and the U. S. Office of Naval Research under the Joint Services Electronics Program by Contracts N00014-67-A-0298-0006, 0005, and 0008.

Division of Engineering and Applied Physics

Harvard University · Cambridge, Massachusetts

lc

ESTIMATION FOR ROTATIONAL PROCESSES WITH ONE DEGREE OF FREEDOM

By

James Ting-Ho Lo and Alan S. Willsky*

Division of Engineering and Applied Physics

Harvard University · Cambridge, Massachusetts

ABSTRACT

A class of bilinear estimation problems involving single-degree-of-freedom rotation is formulated and resolved. Both continuous and discrete time estimation problems are considered. Error criteria, probability distributions, and optimal estimates on the circle are studied. An effective synthesis procedure for continuous time estimation is provided, and a generalization to estimation on arbitrary abelian Lie groups is included. An intrinsic difference between the discrete and continuous problems is discussed, and the complexity of the equations in the discrete time case is analyzed in this setting. Applications of these results to a number of practical problems including FM demodulation and frequency stability are examined.

*Fannie and John Hertz Foundation Fellow, Department of Aeronautics and Astronautics, M. I. T.

TABLE OF CONTENTS

1. Introduction

In the past, most optimal estimation problems have been studied in a vector space setting. While these results lend themselves to simple solutions in linear systems[1,2], and in nonlinear systems with finite dimensional sensor orbits[3], no effective synthesis procedures for optimal estimation have been determined for large classes of nonlinear systems.

It is the purpose of this report to introduce an alternative to the vector space approach in analyzing the properties of nonlinear stochastic processes. We will study random processes on a different type of space, namely a differentiable manifold, which is the natural domain for certain nonlinear problems of practical importance. This approach will be shown to be useful both in analyzing the properties of certain stochastic processes and in deriving recursive optimal estimation equations that are easily implemented (for instance, see the block diagram in Figure 4 and the associated discussion in subsection 3.3).

More specifically, we will concern ourselves with the study of random processes on the circle, S^1, and its extensions to higher dimensions. Topics such as FM demodulation, frequency stability, and single-degree-of-freedom gyroscopic analysis are well-known examples in this framework.

It is appropriate to remark that we will use several distinct representations of the circle interchangeably, depending upon which is most convenient. A point on the unit circle can be represented by either the angle $\theta \in [-\pi, \pi)$ it makes with a fixed reference point on the circle or by the 2×2 orthogonal matrix

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} .$$

Note that the addition of two angles θ_1 and θ_2 modulo 2π corresponds to the multiplication of the two matrices representing the points.

Another representation of S^1 is as the set of complex numbers of length one. Any such number can be uniquely written as $e^{i\theta}$ with $\theta \in [-\pi, \pi)$, and the relationship with the above representations is obvious.

Finally, there exists a natural projection from R^1 to S^1, identified with $[-\pi, \pi)$:

$$x \longmapsto x \bmod 2\pi \ .$$

As Figure 1 indicates, two points x_1 and x_2 are projected onto the same point if and only if they differ by an integral multiple of 2π (that is, $e^{i\theta} = e^{i(\theta + 2n\pi)}$). Thus we divide the real numbers into equivalence classes, $\{x + 2n\pi \mid n \in Z\}$, and to each element of S^1 there corresponds a unique equivalence class, with different points in S^1 corresponding to different equivalence classes. Thus we can represent S^1 by this set of equivalence classes, denoted $R^1/2\pi Z$.

Throughout most of this report the first two representations will be used. However, in Section 5 we will use the complex number representation, and in Section 4 we will make use of the interpretation given by the last representation above.

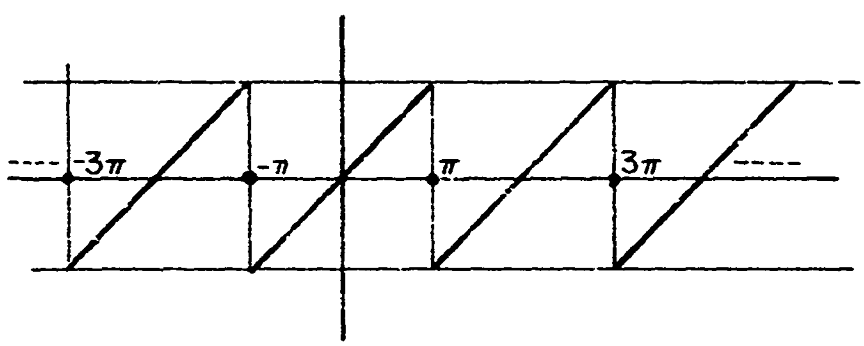

Figure 1. Illustrating the Projection Map

Consider the situation depicted in Figure 2. We have a unit circle in R^2 with a straight line of infinite length tangent to it.

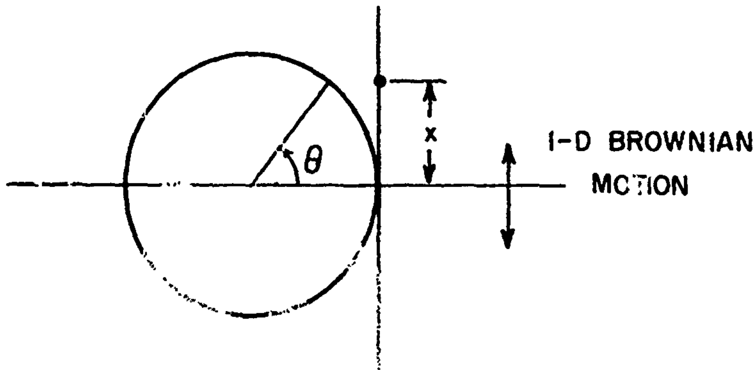

Figure 2

We allow the line to perform a one-dimensional Brownian motion, fix the center of the circle, and require that there be no slipping at the point of tangency. The line induces a rotation of the circle, and, if the line moves a distance x, the circle rotates x radians, and is thus in a position which is $x \bmod 2\pi = \theta$ radians away from its initial position.

The probability density function for θ satisfies the classical heat (diffusion, Fokker-Planck) equation on the circle:

$$\frac{\partial p_\theta}{\partial t} - \frac{1}{2}\frac{\partial^2 p_\theta}{\partial \xi^2} = 0 \tag{1}$$

with the periodicity condition

$$p_\theta(\xi, t) = p_\theta(\xi + 2\pi, t) \tag{2}$$

and initial condition

$$p_\theta(\xi, 0) = \delta(\xi - \eta) \quad . \tag{3}$$

where the initial orientation of the circle is η radians from some reference position. The solution of (1), (2), and (3) is widely known, and is given by the two equivalent ex[illegible]essions

$$p_\theta(\xi, t) = \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{+\infty} e^{-\frac{(\xi + 2n\pi - \eta)^2}{2t}} \tag{4a}$$

$$= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-n^2 t/2} \cos n(\xi - \eta) \quad . \tag{4b}$$

The density in (4) will be called the <u>folded normal density</u>. We give it this name for the following reason: if x is a normal random variable with mean η and variance γ, and if we let $\theta = x \bmod 2\pi$, then the density for θ, p_θ, is given by

$$p_\theta(\xi) = \frac{1}{\sqrt{2\pi\gamma}} \sum_{n=-\infty}^{+\infty} e^{-\frac{(\xi + 2n\pi - \eta)^2}{2\gamma}} \triangleq F(\xi; \eta, \gamma) \quad .$$

Levy[4], and Perrin[5] have done extensive work with this density.

Using this concept of "wrapping" a random process around the circle, we formulate the mathematical model of an observation process that can be described by a bilinear matrix Ito stochastic equation. Let m be a random process on R^1, and define z by

$$dz(t) = m(t)dt + dw(t)$$

where w is a Brownian motion process independent of m. Consider the associated process

$$\theta(t) = z(t) \bmod 2\pi \quad .$$

Since knowledge of $\sin\theta$ and $\cos\theta$ is equivalent to knowledge of θ, we wish to find an equation for

$$Z(t) = \begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix} \quad .$$

As will be shown

$$dZ(t) = Z(t) \begin{bmatrix} -\frac{1}{2}dt & m(t)dt+dw(t) \\ -m(t)dt-dw(t) & -\frac{1}{2}dt \end{bmatrix} , \qquad (5)$$

where the $-\frac{1}{2}dt$ terms are the second order correction terms given by Ito stochastic calculus[6,7]. These terms are precisely what is needed to insure (in the Ito sense) that $Z(t)$ remains an orthogonal matrix.

If we assume $z(0) = 0$, we can write

$$z(t) = \int_0^t m(s)ds + w(t) \quad ,$$

and then

$$Z(t) = \begin{bmatrix} \cos w(t) & \sin w(t) \\ -\sin w(t) & \cos w(t) \end{bmatrix} \begin{bmatrix} \cos[\int_0^t m(s)ds] & \sin[\int_0^t m(s)ds] \\ -\sin[\int_0^t m(s)ds] & \cos[\int_0^t m(s)ds] \end{bmatrix} , \qquad (6)$$

and, in this form, we see that the disturbance is multiplicative in nature.

In this report we will examine multiplicative noise problems such as this and will derive estimation equations for them. In Section 2 we will examine various error criteria for the optimal estimation of random

variables on the circle. Section 3 deals with continuous time estimation of a class of stochastic processes on the circle, and Section 4 discusses the discrete time problem. Applications of this theory to AM and FM demodulation, optical communication, frequency stability, and estimation of the orientation of a spinning body are discussed in Section 5. In addition, an appendix is included, in which the relationship between the discrete and continuous time problems of Sections 3 and 4 is discussed.

2. Error Criteria and Optimal Estimates

In the following sections, we will study the properties of certain stochastic processes on the circle and will derive equations for probability distributions conditioned on observations. The question of optimal estimation will be of central importance in Sections 3 and 4. Thus it became necessary to study how one uses the knowledge of the probability distribution of the quantity to be estimated to choose an estimate that gives the "best" performance, as measured by some pre-determined figure of merit.

In this section, we will present a number of results on the optimal estimation of random variables taking values on the circle. We assume that we are given a random variable θ taking on values in $[-\pi, \pi)$, with probability density $p(\theta)$, which is assumed to be periodic with period 2π. Also, we assume that we have an error function ϕ, also periodic with period 2π, and we wish to choose $\hat{\theta}$ to minimize

$$\mathscr{E}(\phi(\theta - \hat{\theta})) = \int_{-\pi}^{\pi} \phi(\theta - \hat{\theta})p(\theta)d\theta \quad .$$

This is precisely the S^1 analog of the vector space optimal estimation problem[6].

The motivation throughout this section is to provide simple methods for computing the minimum of the cost criterion, $\mathscr{E}(\phi(\theta - \hat{\theta}))$, and the value $\hat{\theta}$ that achieves this minimum. In this light, a number of special cases (i.e. particular families of densities and error functions) are considered in detail.

The first subsection presents a basic result, analogous to Sherman's results[10, 11], on optimal estimates for a large class of error criteria, but

for the rather special case of unimodal probability density functions. However, it is shown that the important folded normal density falls into this class.

The second subsection deals with the more general estimation problem, in which the density need not be unimodal and the error function may have a more general shape. Fourier series is the basic tool of this section. The third subsection contains detailed analysis for the special cases of the folded normal density and a linear combination of folded normal densities.

2.1 Symmetric Criteria and Unimodal Distributions

We define the standard distance function (Riemannian metric) on the circle -- i.e. the distance, ρ, between two points on the circle is the arc length of the shortest path (geodesic line) joining them. If we restrict θ_1 and θ_2 to take values in the range $[-\pi, \pi)$, we have

$$\rho(\theta_1, \theta_2) = \min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|) \quad .$$

The class of error criteria we wish to consider is the class of symmetric, nondecreasing cost functions -- i.e. functions $\phi: S^1 \longrightarrow R$ which satisfy

$$\begin{aligned} &0 \leq \phi(\theta) = \phi(-\theta) \\ &0 \leq \rho(\theta_1, 0) \leq \rho(\theta_2, 0) \Longrightarrow \phi(\theta_1) \leq \phi(\theta_2) \quad . \end{aligned} \tag{7}$$

Some examples of cost criteria satisfying (7) are $\rho(\theta) \stackrel{\Delta}{=} \rho(\theta, 0)$, $(1 - \cos\theta)$, $\rho(\theta)^2$, $(1 - \cos\theta)^2$. We also wish to consider the special class of unimodal, mode-symmetric probability density functions -- i.e. density functions of

the form $p: S^1 \to [0, \infty)$ with a unique maximum at η, such that

$$p(\eta + \theta) = p(\eta - \theta) \quad . \qquad \forall \theta$$

As the following theorem demonstrates, under these conditions the mode of the density is the optimal estimate.

Theorem 1: Given an error function ϕ that satisfies (7) and a unimodal, mode-symmetric probability density function p, then

$$\mathscr{E}(\phi(\theta - \eta)) \leq \mathscr{E}(\phi(\theta - a)) \qquad \forall a$$

where p has its maximum at η.

Proof: The theorem follows immediately from results on similarly ordered functions and the rearrangement inequalities. The basic result for real valued functions defined on R^1 is contained in Hardy, Littlewood and Polya[8] (thm. 378) and Szego and Polya[9] (p. 183). The result for S^1 is obtained by making only minor changes in these proofs. ■

We remark that from the symmetry of the problem, ϕ has its global maximum at π and p has its global minimum at $\eta + \pi$. Thus

$$\mathscr{E}(\phi(\theta - \eta + \pi)) \geq \mathscr{E}(\phi(\theta - a)) \qquad \forall a .$$

It should be noted that Theorem 1 is the S^1 analog of a result of Sherman[10,11]. Note that the same result is true if a probability density doesn't exist, but the probability measure is unimodal at, and symmetric about some point η. Here we define these concepts as follows: let θ be a random variable on S^1 and define the distribution function $F: [-\pi, \pi] \to [0, 1]$ by

$$F(a) = Pr(\theta \in [-\pi, a]) \quad .$$

Then F is unimodal at, and symmetric about 0 if it is convex for $\alpha \in [-\pi, 0)$, and if

$$F(\alpha) = 1 - F(-\alpha)$$

at each continuity point of F (see ref. 10).

In the continuous time problem discussed in Section 3 and the discrete time problem of Section 4, the folded normal distribution will play an important role, and for this density we have the following result which shows that Theorem 1 holds for the folded normal density.

Theorem 2: The folded normal density

$$\begin{aligned} F(\theta; \eta, \gamma) &= \frac{1}{\sqrt{2\pi\gamma}} \sum_{n=-\infty}^{+\infty} e^{-\frac{(\theta+2n\pi-\eta)^2}{2\gamma}} \\ &= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-n^2\gamma/2} \cos n(\theta - \eta) \end{aligned} \tag{8}$$

is unimodal with mode at $\theta = \eta$ and is symmetric about η.

Proof: Since $\cos \leq 1$, the second form of F in (8) yields

$$F(\theta; \eta, \gamma) \leq \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-n^2\gamma/2} = F(\eta; \eta, \gamma) \quad .$$

Thus F has its global maximum at $\theta = \eta$.

Since $F(\theta, \eta, \gamma) = F(\theta - \eta; 0, \gamma)$, we need only show that $F(\theta; 0, \gamma)$ is symmetric about 0 and monotone decreasing as $\rho(\theta, 0)$ increases. Symmetry is obvious $(\cos n\theta = \cos n(-\theta))$, and monotonicity will follow if we can show

$$\frac{\partial F}{\partial \theta}(\theta, 0, \gamma) < 0 \qquad \theta \in (0, \pi) \tag{9a}$$

$$\frac{\partial F}{\partial \theta}(\theta;\ 0, \gamma) > 0 \qquad \theta \in (-\pi, 0) \tag{9b}$$

We now remark that the properties of $F(\theta; 0, \gamma)$ have been studied extensively, since it is a theta function. See refs. 12 and 13 for discussions of some properties of theta functions. Using the notation of ref. 12, pp. 2, 42, we have

$$\begin{aligned} F(\theta; 0, \gamma) &= \frac{1}{2\pi}\ \theta_4\left(\frac{\theta+\pi}{2}, \frac{i\gamma}{2\pi}\right) \\ &= k \prod_{n=1}^{\infty} (1 + 2q^{2n-1}\cos\theta + q^{4n-2}) , \end{aligned} \tag{10}$$

where

$$q = e^{-\gamma/2}$$

and

$$k = \frac{1}{2\pi} \prod_{n=1}^{\infty} (1 - q^{2n}) .$$

Using the fact that $F > 0$ and the form of F given by (10), we have

$$\frac{\frac{\partial F}{\partial \theta}(\theta; 0, \gamma)}{F(\theta; 0, \gamma)} = -\left[\sum_{n=1}^{\infty} \frac{2q^{2n-1}}{(1 + 2q^{2n-1}\cos\theta + q^{4n-2})}\right] \sin\theta . \tag{11}$$

It is easily seen that the term in square brackets on the right hand side of (11) is positive for all values of θ and thus (9) is correct. ■

Some work along these lines has been done by Perrin[5]. See ref. 15 for discussions of other relevant properties of theta functions, hypergeometric

functions, Legendre polynomials, and Tchebycheff polynomials.

Note that the symmetry requirements of Theorem 1 are necessary. For instance, if ϕ is not symmetric, the mode of the density need not be the optimal estimate even if all the other assumptions of Theorem 1 do hold. As an example, consider the function $\phi: S^1 \longrightarrow R$

$$\phi(\theta) = \begin{cases} \theta & 0 \leq \theta \leq \pi \\ \dfrac{\theta^2}{\pi} & -\pi \leq \theta \leq 0 \end{cases} .$$

Suppose our distribution is the folded normal centered at 0. Then it can be shown that the mode, 0, is not the optimal estimate.

2.2 Optimal Estimation Using Fourier Series

If we do not have a unimodal distribution or symmetric cost criteria that increases away from 0, Theorem 1 doesn't apply, but, with the aid of Fourier series, we can still do some useful analysis. We assume that our probability density is given in Fourier series form

$$p(\theta) = \frac{1}{2\pi} + \sum_{n=1}^{\infty} a_n \sin n\theta + b_n \cos n\theta$$

as is our error function

$$\phi(\theta) = d_o + \sum_{n=1}^{\infty} c_n \sin n\theta + d_n \cos n\theta \quad .$$

Our problem is to choose $\hat{\theta}$ to minimize $E(\phi(\theta - \hat{\theta}))$. A simple computation yields

$$\mathscr{E}(\phi(\theta - \hat{\theta})) = d_o + \pi \sum_{n=1}^{\infty} \Big\{ a_n(c_n \cos n\hat{\theta} + d_n \sin n\hat{\theta})$$

$$+ b_n(d_n \cos n\hat{\theta} - c_n \sin n\hat{\theta}) \Big\} \quad . \qquad (12)$$

Thus, necessary conditions for a local minimum are

$$\frac{d}{d\hat{\theta}} \mathscr{E}(\phi(\theta - \hat{\theta})) = 0 \implies$$

$$\sum_{n=1}^{\infty} \Big\{ na_n[d_n \cos n\hat{\theta} - c_n \sin n\hat{\theta}] - nb_n[d_n \sin n\hat{\theta} + c_n \cos n\hat{\theta}] \Big\} = 0 \qquad (13)$$

$$\frac{d^2}{d\hat{\theta}^2} \mathscr{E}(\phi(\theta - \hat{\theta})) \geq 0 \implies$$

$$\sum_{n=1}^{\infty} -n^2 a_n[d_n \sin n\hat{\theta} + c_n \cos n\hat{\theta}] + n^2 b_n[c_n \sin n\hat{\theta} - d_n \cos n\hat{\theta}] \geq 0 \quad . \qquad (14)$$

Solutions of (13) and (14) are candidates for the optimal estimate.

Explicit solution of (13) and (14) is possible only for certain error functions. For example, suppose we consider the function

$$\phi_1(\theta) = 1 - \cos \theta \quad .$$

Then $d_o = 1$, $d_1 = -1$, and all other Fourier coefficients are 0. Then

$$\mathscr{E}(\phi_1(\theta - \hat{\theta})) = 1 - \pi(a_1 \sin \hat{\theta} + b_1 \cos \hat{\theta}) \qquad (15)$$

and equations (13) and (14) become

$$a_1 \cos \hat{\theta} - b_1 \sin \hat{\theta} = 0 \qquad (16)$$

$$a_1 \sin \hat{\theta} + b_1 \cos \hat{\theta} \geq 0 \quad . \tag{17}$$

If $a_1 = b_1 = 0$, $\mathscr{E}(\phi_1(\theta - \hat{\theta}))$ is independent of $\hat{\theta}$. In any other case, there are two inequivalent solutions to (16), where two solutions are considered equivalent if they differ by a multiple of 2π. The two solutions are

$$\hat{\theta} = \tan^{-1}(a_1/b_1), \ \tan^{-1}(a_1/b_1) + \pi$$

where $\tan^{-1}:[-\infty, \infty] \longrightarrow [-\pi/2, \pi/2]$. Examination of (15) and (17) yields a method for choosing the proper solution:

$a_1 \geq 0, \ b_1 \geq 0 \implies$ choose solution in <u>first</u> quadrant

$a_1 \geq 0, \ b_1 < 0 \implies$ choose solution in <u>second</u> quadrant

$a_1 < 0, \ b_1 < 0 \implies$ choose solution in <u>third</u> quadrant

$a_1 < 0, \ b_1 \geq 0 \implies$ choose solution in <u>fourth</u> quadrant.

With . e choices, it is easy to see that

$$\sin \hat{\theta}_0 = \frac{a_1}{\sqrt{a_1^2 + b_1^2}}$$

$$\cos \hat{\theta}_0 = \frac{b_1}{\sqrt{a_1^2 + b_1^2}}$$

and

$$\mathscr{E}(\phi(\theta - \hat{\theta}_0)) = 1 - \pi \sqrt{a_1^2 + b_1^2} \quad , \tag{18}$$

where $\hat{\theta}_0$ is the optimal value.

Thus, in this case, we can explicitly solve the estimation problem in terms of the first mode Fourier coefficients. Note that the higher modes

play no role in this particular case, but also note that ϕ_1 has some motivation from standard vector space theory, in that for small values of θ,

$$\phi_1(\theta) = 1 - \cos\theta \approx \frac{1}{2}\theta^2 \quad .$$

Another possible error function, one that involves the first and second modes of the density, is

$$\phi_2(\theta) = (1 - \cos\theta)^2 = \frac{3}{2} - 2\cos\theta + \frac{1}{2}\cos 2\theta \quad .$$

Using the same type of approach as before, one can reduce the problem of finding the optimal estimate to the solution of a quartic polynomial equation and the calculation of several functions -- a procedure that can be done easily by computer. However, the complexity, even when we just add in the second mode, is such that no closed form for the optimal error in terms of the Fourier coefficients is available.

As can be seen, the error analysis becomes increasingly more difficult as the number of nonzero Fourier coefficients increases. For example, direct application of these ideas if $\phi = \rho$ or ρ^2, where ρ is the Riemannian metric on S^1 (actually $\rho(\theta) \triangleq \rho(\theta, 0)$), yields extremely complicated equations. However, the $\frac{1}{n^2}$ behavior of the Fourier coefficients for these two examples suggests truncating the series and applying techniques such as those used in the analysis for $(1 - \cos\theta)$ and $(1 - \cos\theta)^2$.

However, for these special functions we can use a different method in trying to find the optimal estimate. Consider the function ρ^2. We have the equation

$$\rho^2(\theta) = \theta^2 \qquad -\pi \le \theta \le \pi \quad .$$

Thus, if our probability density is $p(\theta)$,

$$\mathscr{E}(\rho^2(\theta - \hat{\theta})) = \int_{-\pi+\hat{\theta}}^{\pi+\hat{\theta}} (\theta - \hat{\theta})^2 \, p(\theta) \, d\theta \quad .$$

Using Leibnitz's rule and the periodicity of p, we have the following necessary conditions for optimality

$$\frac{d}{d\hat{\theta}} \mathscr{E}(\rho^2(\theta - \hat{\theta})) = 2\hat{\theta} - 2 \int_{-\pi+\hat{\theta}}^{\pi-\hat{\theta}} \theta \, p(\theta) \, d\theta = 0 \tag{19}$$

and

$$\frac{d^2}{d\hat{\theta}^2} \mathscr{E}(\rho^2(\theta - \hat{\theta})) = 2 - 4\pi \, p(\hat{\theta} + \pi) \geq 0 \quad . \tag{20}$$

Equations (19) and (20) offer an alternate method for solving for θ_0. Note that equation (19) resembles the necessary condition for the least squares estimate on R^1. In that case

$$\hat{x}_0 = \mathscr{E}(x) = \int_{-\infty}^{+\infty} x \, p(x) dy \ ,$$

where $p(x)$ is the density function. However, in this case, essentially because of the topological difference between S^1 and R^1, the integral

$$\int_{-\pi+\hat{\theta}}^{\pi+\hat{\theta}} \theta \, p(\theta) d\theta$$

is not independent of $\hat{\theta}$, and thus cannot be called $\mathscr{E}(\theta)$.

2.3 The Folded Normal Density and Its Linear Combinations

As will be seen in the next two sections, two types of probability densities are of great importance. The first of these is the folded normal density, $F(\theta; \eta, \gamma)$, and the second is a linear combination of such densities

$$p(\theta) = \sum_{n=1}^{\infty} c_n F(\theta; \eta, \gamma_n) \tag{21}$$

$$\sum_{n=1}^{\infty} c_n = 1, \qquad \gamma_n > 0 \quad .$$

It should be noted that it has been shown[3] that the set of densities given by (21) with only finitely many nonzero c_n's is dense in $L^1(-\pi, \pi)$, and this is still true if all the γ_n's are equal to some fixed γ. In this section we do not require that only finitely many c_n's be unequal to zero. The reason for this will be seen in Section 4.

For the case where our density $p(\theta)$ is a single folded normal density, $F(\theta; \eta, \gamma)$, we know that the optimal estimate for any function ϕ satisfying (7) is the mode, η. However, for this special density, we can say a great deal more. Let us consider a more general class of error functions. We remove the symmetry requirement but still require that ϕ be increasing on $[0, \pi]$ and decreasing on $[-\pi, 0]$. For such a ϕ, the mode η need not be the optimal estimate, however for this discussion we will take it as our estimate. The following theorem reveals an important property of the error $\mathscr{E}(\phi(\theta - \eta))$.

<u>Theorem 3:</u> For ϕ satisfying the above requirement, and $p(\theta) = F(\theta; \eta, \gamma)$,

$\mathcal{E}(\phi(\theta - \eta))$ is an increasing function of the variance, γ -- that is

$$\frac{d}{d\gamma}\,\mathcal{E}(\phi(\theta - \eta)) \geq 0 \tag{22}$$

Proof: Writing

$$\phi(\theta) = d_o + \sum_{n=1}^{\infty} c_n \sin n\theta + d_n \cos n\theta$$

and using the results on Fourier series analysis,

$$\mathcal{E}(\phi(\theta - \eta)) = d_o + \sum_{n=1}^{\infty} d_n\, e^{-n^2\gamma/2} , \tag{23}$$

but we get the same error if we compute $\mathcal{E}(\psi(\theta - \eta))$, where ψ is the function satisfying (7) defined by

$$\psi(\theta) = \frac{1}{2}(\phi(\theta) + \phi(-\theta)) .$$

Thus, it is enough to prove the theorem for ϕ satisfying (7). In this case η is the optimal estimate and

$$\begin{aligned}\mathcal{E}(\phi(\theta - \eta)) &= \int \phi(\theta - \eta) F(\theta, \eta, \gamma) d\theta \\ &= \int_{-\pi}^{\pi} \phi(\theta) F(\theta; 0, \gamma) d\theta \\ &= 2 \int_0^{\pi} \phi(\theta) F(\theta; 0, \gamma) d\theta .\end{aligned}$$

Then, (22) will hold if

$$\int_0^{\pi} \phi(\theta) \frac{\partial}{\partial\gamma} F(\theta; 0, \gamma)\, d\theta \geq 0 .$$

Suppose we can show that there exists $\theta_0 \in [0,\pi]$ such that

$$\frac{\partial}{\partial\gamma} F(\theta;0,\gamma) < 0 \qquad \theta \in [0,\theta_0)$$

$$\frac{\partial}{\partial\gamma} F(\theta_0;0,\gamma) = 0$$

$$\frac{\partial}{\partial\gamma} F(\theta;0,\gamma) > 0 \qquad \theta \in (\theta_0,\pi] \quad .$$

Then, since

$$\phi(\theta) \le \phi(\theta_0) \qquad \theta \in [0,\theta_0]$$

$$\phi(\theta) \ge \phi(\theta_0) \qquad \theta \in [\theta_0,\pi] \quad ,$$

we have

$$\int_0^\pi \phi(\theta) \frac{\partial}{\partial\gamma} F(\theta;0,\gamma)d\theta \ge \phi(\theta_0) \frac{d}{d\gamma} \int_0^\pi F(\theta;0,\gamma)d\theta$$

$$= \phi(\theta_0) \frac{d}{d\gamma} \left(\frac{1}{2}\right) = 0 \quad ,$$

and we get a strict inequality if ϕ is not a constant.

Now it is easy to see that

$$\frac{\partial}{\partial\gamma} F(\theta;0,\gamma) = \frac{1}{2} \frac{\partial^2}{\partial\theta^2} F(\theta;0,\gamma)$$

and the theorem will be proved once we prove the following lemma, which yields more information about the shape of the folded normal density.

Lemma 1: For an arbitrary but fixed value of $\gamma > 0$, there exists $\theta_0 \in [0,\pi]$ such that

$$\frac{\partial^2}{\partial\theta^2} F(\theta; 0, \gamma) < 0 \qquad \theta \in [0, \theta_0)$$

$$\frac{\partial^2}{\partial\theta^2} F(\theta_0; 0, \gamma) = 0$$

$$\frac{\partial^2}{\partial\theta^2} F(\theta; 0, \gamma) > 0 \qquad \theta \in (\theta_0, \pi]$$

that is, F has a unique inflection point (at θ_0) on $[0, \pi]$.

<u>Proof</u>: We use the form of $F(\theta; 0, \gamma)$ given in equation (10). We compute

$$\frac{\frac{\partial^2}{\partial\theta^2} F}{F} = - A \cos\theta + B \sin^2\theta$$

$$A = \sum_{n=1}^{\infty} \frac{2q^{2n-1}}{(1+2q^{2n-1}\cos\theta + q^{4n-2})}$$

$$B = \sum_{n \neq m} \frac{4q^{2(n+m-1)}}{(1+2q^{2n-1}\cos\theta + q^{4n-2})(1+2q^{2m-2}\cos\theta + q^{4m-2})}$$

and then a simple computation yields

$$\frac{\partial^2}{\partial\theta^2} F(0; 0, \gamma) < 0$$

$$\frac{\partial^2}{\partial\theta^2} F(\theta; 0, \gamma) > 0 \qquad \forall\, \theta \in [\tfrac{\pi}{2}, \pi]$$

and

$$\frac{\partial}{\partial\theta}\left(\frac{\frac{\partial^2}{\partial\theta^2}F}{F}\right)(\theta;0,\gamma) > 0 \qquad \forall\,\theta \in (0, \tfrac{\pi}{2})$$

These inequalities imply that, there is a $\theta_0 \in (0, \frac{\pi}{2})$ such that $\frac{\partial^2}{\partial\theta^2} F(\theta_0; 0, \gamma) = 0$. Then for $\theta_0 < \theta_1 \leq \pi/2$

$$\frac{\frac{\partial^2}{\partial\theta^2} F(\theta_1; 0, \gamma)}{F(\theta_1; 0, \gamma)} > \frac{\frac{\partial^2}{\partial\theta^2} F(\theta_0; 0, \gamma)}{F(\theta_0; 0, \gamma)} = 0$$

or

$$\frac{\partial^2}{\partial\theta^2} F(\theta_1; 0, \gamma) > 0$$

and the lemma and the theorem are proved. ■

Note that by symmetry we have that F has a unique inflection point at $-\theta_0$ on the interval $[-\pi, 0]$.

Theorem 3 tells us that the intuitive notion that we "have more accurate information" for smaller values of γ can be made precise. Also, this theorem implies another result, which is the S^1 analog of a problem treated by J. L. Brown[14]. The problem treated in ref. 14 is that of finding the optimal linear filter minimizing an asymmetric error criteria on R^1 that decreases on $(-\infty, 0]$ and increases on $[0, \infty)$. The result is that the optimal linear filter is the minimum variance filter, and the proof essentially consists of showing that the error is an increasing function of the variance. Theorem 3 clearly implies an S^1 analog of this result.

Some examples of cost criteria satisfying (7) and the associated optimal costs when the density is folded-normal will be given in Section 3.

For the case in which $p(\theta)$ is given by (21), the situation is somewhat different and much more complicated, since we no longer have a unimodal probability density. For this case, we will examine the optimal estimation problem for two error functions, $1-\cos\theta$ and $\rho^2(\theta)$.

As discussed in subsection 2.2, in trying to minimize $E(1-\cos(\theta-\hat{\theta}))$ with respect to $\hat{\theta}$, we need only know the lowest mode Fourier coefficients, a_1 and b_1. In this case

$$a_1 = \frac{1}{\pi} \sum_{n=1}^{\infty} c_n e^{-\gamma_n/2} \sin \eta_n$$

$$b_1 = \frac{1}{\pi} \sum_{n=1}^{\infty} c_n e^{-\gamma_n/2} \cos \eta_n$$

and (assuming a_1 and b_1 are not both zero) the optimal estimate $\hat{\theta}_0$ is either $\tan^{-1} a_1/b_1$ or $\tan^{-1} a_1/b_1 + \pi$, depending upon the signs of a_1 and b_1. In any case, the optimal cost is given by

$$E(1-\cos(\theta-\hat{\theta}_0)) = 1 - \left\{ \left[\sum_{n=1}^{\infty} c_n e^{-\gamma_n/2} \sin \eta_n \right]^2 + \left[\sum_{n=1}^{\infty} c_n e^{-\gamma_n/2} \cos \eta_n \right]^2 \right\}^{1/2} \tag{24}$$

In general, this optimal error is <u>not</u> an increasing function of each of the variances γ_n, individually. However, if all of the variances equal some value γ, it is easy to see that the optimal error <u>is</u> an increasing function of γ.

In the case of $\rho^2(\theta)$, we recall from subsection 2.2 that it was necessary to evaluate

$$\int_{-\pi+\alpha}^{\pi+\alpha} \theta\ p(\theta)\ d\theta$$

as a function of α. For p a folded normal density, $F(\theta;\eta,\gamma)$, we have

$$\int_{-\pi+\alpha}^{\pi+\alpha} \theta\ p(\theta) d\theta = \eta - \sum_{k=-\infty}^{+\infty} 2k\pi \int_{(2k-1)\pi}^{(2k+1)\pi} N(\theta;\eta-\alpha,\gamma) d\theta \tag{25}$$

where N is the normal density. The second term on the right-hand side of (25) involves various values of the error function, erf, and can be tabulated as a function of $\eta-\alpha$ and γ. Then, if we call this term $g(\eta-\alpha,\gamma)$, in the case where $p(\theta)$ is given by (21), necessary conditions for the optimal estimate are

$$\hat{\theta}_0 = \sum_{n=1}^{\infty} c_n [\eta_n - g(\eta_n - \hat{\theta}_0, \gamma_n)] \tag{26}$$

$$1 - 2\pi \sum_{n=1}^{\infty} c_n F(\hat{\theta}_0 + \pi; \eta_n, \gamma_n) \geq 0 \tag{27}$$

There does not appear to be a simple formula for the optimal cost, nor is it clear whether or not the optimal cost is a monotone increasing functions of the γ_n or of γ, in the case where all of the $\gamma_n = \gamma$.

3. Continuous Time Estimation

A signal process and an observation process, taking values on S^1, will be formulated in terms of bilinear Ito matrix differential equations. The conditional probability distribution of the signal, given observations over a certain period of time, will be evaluated. Recursive computational schemes for optimal estimation (filtering, smoothing, and prediction), with respect to the error criteria defined in the previous section, will be derived. In fact it will be shown that optimal estimates on S^1 can be obtained recursively by the use of an ordinary vector space estimator together with a nonlinear preprocessor and a nonlinear postprocessor, as illustrated in Fig. 4. Multichannel estimation on abelian Lie groups will be examined. Examples illustrating the optimal estimation procedure are given at the end of this section.

The circle group, S^1, can be identified as the multiplication group of 2×2 orthogonal matrices of determinant +1. Any element of this group has the form

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

and, for θ near zero, we have the first order approximation

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \approx \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \theta \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

The matrix

$$R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

is called the <u>infinitesimal rotation</u>, and we have

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \exp R\theta$$

For those familiar with the theory of Lie groups, S^1 is a one dimensional abelian Lie group, with the 2×2 orthogonal matrices a representation of the group. The infinitesimal rotation R forms a basis for the Lie algebra, $L(S^1)$, of S^1. The Lie algebra and Lie group are related by the exponential map

$$\exp(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!} \qquad A \in L(S^1)$$

and the logarithm map

$$\log(B) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(B-I)^n}{n} \qquad B \in S^1, |B-I| < 1$$

3.1 Signal Processes and Observation Processes

It has been shown [21, p. 269] that the circular Brownian motion on S^1 can be constructed by taking the projection modulo 2π of the standard 1-dimensional Brownian motion onto the unit circle S^1. This method will now be used to construct a continuous signal process on S^1 and to formulate the mathematical model of a sensor (an observation process) to be used in this report.

We will adopt the following notation

$(\Omega, \mathcal{A}, P)$ = a probability space

s - a positive real number

C_1^s = the family of real-valued continuous functions, a, on $[0, s]$ such that $a(0) = 0$

$\mathscr{B}_1^s$ = the Borel σ-field of C_1^s

C_2^s = the family of 2×2 orthogonal-matrix-valued continuous functions, A, on $[0, s]$ such that $A(0) = I$, the identity matrix

$\mathscr{B}_2^s$ = the Borel σ-field of C_2^s

Lower case letters denote elements in C_1^s and upper case letters denote elements in C_2^s.

Let $J: C_1^s \longrightarrow C_2^s$ be defined by

$$(J(a))(t) = \exp(a(t)R) = \begin{bmatrix} \cos a(t) & \sin a(t) \\ -\sin a(t) & \cos a(t) \end{bmatrix} \tag{28}$$

for $a \in C_1^s$ and $t \in [0, s]$. It is easily seen that J is $\mathscr{B}_1^s$-measurable and bijective. This bijective operator will play a key role in this section. Intuitively, J can best be illustrated by Fig. 3. A point on the unit circle, S^1, can be represented by either the angle $\theta \in [-\pi, \pi)$ it makes with a fixed radial axis or the 2×2 orthogonal matrix $\exp(R\theta)$. Therefore, in the first representation, C_2^s is the family of piecewise continuous functions $\theta(t)$. such that at any point of discontinuity the right hand limit of θ is $\pm\pi$, while the left-hand limit is $\mp\pi$ (see Fig. 3).

Each continuous curve a(t) on R^1 gives rise to <u>one</u> <u>and</u> <u>only</u> <u>one</u> piecewise continuous curve $\theta(t)$ lying between π and $-\pi$, of which the continuous segments are obtained by translating the corresponding segments of a(t) an integral number of multiples of 2π (see Fig. 3).

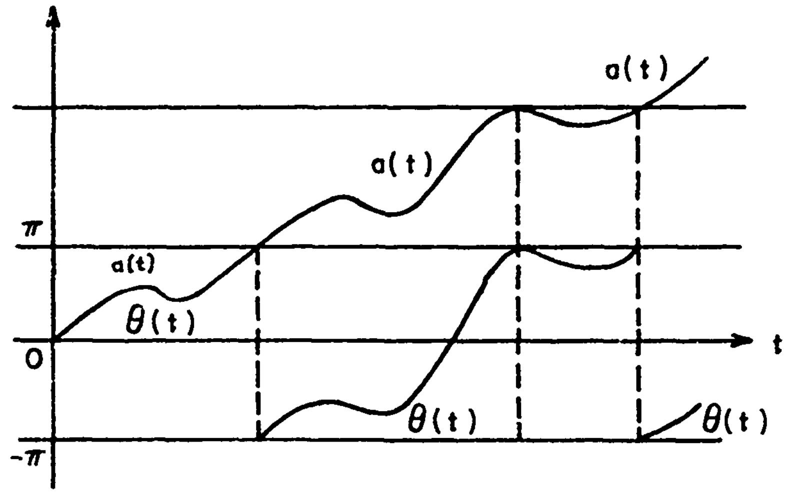

Figure 3

Conversely, each piecewise continuous curve in C_2 gives rise to <u>one and only one</u> continuous curve taking values on R^1 which is obtained simply by piecing the continuous segments together. This intuitive observation illustrates the bijective property of the operator J. Thus a continuous random signal process on S^1 which is described by an $\mathscr{A}$-measurable function $X:\Omega \longrightarrow C_2^s$ corresponds to a continuous random signal process on R^1 which is described by an $\mathscr{A}$-measurable function $x:\Omega \longrightarrow C_1^s$ such that

$$X(t) = (J(x))(t) , \qquad t \in [0, s] . \tag{29}$$

We now define a random process $z:\Omega \longrightarrow C_1^s$ by the K. Ito random differential equation,

$$dz(t) = m(x(t), t)\, dt + q^{1/2}\, d\omega(t), \quad z(0) = 0, \tag{30}$$

where $m: R_1 \times R_1 \to R_1$ is Borel-measurable, $q: R_1 \to R_1$ is positive and measurable and w is the standard Brownian motion on $(\Omega, \mathscr{A}, P)$, independent of x. Let $Z: \Omega \to C_2^s$ be defined by

$$Z(t) = (J(z))(t) . \tag{31}$$

Applying the Ito differentiation rule, we obtain the following Ito matrix differential equation:

$$dZ(t) = Z(t) \begin{bmatrix} -\frac{q(t)}{2} & m(t) \\ - m(t) & -\frac{q(t)}{2} \end{bmatrix} dt + Z(t) \begin{bmatrix} 0 & dw(t) \\ -dw(t) & 0 \end{bmatrix} \tag{32}$$

$$Z(0) = I,$$

where $m(t) \triangleq m(x(t), t)$ and the diagonal terms $\frac{q(t)}{2}$ are the second order correction terms which keep Z on the circle. This equation is the mathematical model of the sensor to be used. We note that the input, x(t) to the sensor is not the dynamical state X(t) of the rotational signal process on the circle, but rather the angle the rotational process has swept.

The physical motivation for this sensor model comes from the fact that in observing a rotational process (for instance a gyroscope recording rotation about a fixed axis) our measurement contains information on the total rotation, x(t), not just the orientation, X(t). In some applications, such as the gyro problem mentioned above, we wish to extract knowledge of orientation from knowledge of rotation, so it is proper to regard X(t) as the signal process. However, in other applications, such as FM demodulation, our interest centers on the x process, and in these cases, we

may regard x as the signal.

3.2 Conditional Probability Distributions

In this subsection, we will derive equations for the conditional probability distribution of the signal process given observations over some time period. The approach of this section is measure-theoretic in nature, and the major results are summarized in the statements of Lemma 2, Theorem 4, and its two corollaries.

Let us denote $\{z(\tau), \tau \in [0,t]\}$ and $\{Z(\tau), \tau \in [0,t]\}$ by z^t and Z^t, respectively. We note that $Z^t = J(z^t)$. Since J is bijective from C_1^t to C_2^t. the σ-subfield of $\mathscr{A}$ generated by z^t is the same as that generated by Z^t. In other words, the information carried by z^t and Z^t is the same. That σ-subfield will be denoted by $\mathscr{A}_z^t$. The σ-subfield of $\mathscr{A}$ which is generated by $X_\lambda = X(\lambda)$ (the subscripts λ, s, t denote that the processes are evaluated at these times.) will be denoted by $\mathscr{A}_x$.

Let P_{xz} be the conditional probability measure on $(\Omega, \mathscr{A}_x)$ given $\mathscr{A}_z^t$, defined by $P_{xz}(A, \omega_2) = P(A|\mathscr{A}_z^t)(\omega_2)$. for $A \in \mathscr{A}_x$, $\omega_2 \in \Omega$. Let P_{zx} be the conditional probability measure on $(\Omega, \mathscr{A}_z^t)$ given $\mathscr{A}_x$, defined by $P_{zx}(B, \omega_1) = P(B|\mathscr{A}_x)(\omega_1)$, for $B \in \mathscr{A}_z^t$, $\omega_1 \in \Omega$. The restrictions of P to $\mathscr{A}_z^t$ and $\mathscr{A}_x$ are denoted by P_z and P_x, respectively. Let μ_z and μ_w be the induced measures on $(C_1^t, \mathscr{B}_1^t)$ by z^t and w^t, respectively. Define the conditional measure μ_{zx} on $(C_1^t, \mathscr{B}_1^t)$, given X_λ, by $\mu_{zx}(B, \omega_1) = P(z^{-1}(B)|\mathscr{A}_x)(\omega_1)$, for $B \in \mathscr{B}_1^t$, $\omega_1 \in \Omega$.

It is known (ref. 15) that $\mu_{zx} \approx \mu_w \approx \mu_z$ where $\approx$ denotes equivalence of measures, and

$$\frac{d\mu_{zx}}{d\mu_w}(\xi^t, A_\lambda) = \mathscr{E}_x[\theta^t | X_\lambda = A_\lambda] \tag{33}$$

$$\frac{d\mu_z}{d\mu_w}(\xi^t) = \mathscr{E}_x[\theta^t] \tag{34}$$

where $\mathscr{E}_x$ means taking the average over x and

$$\theta^t = \exp\left(-\frac{1}{2}\int_0^t \frac{m^2}{q}(\tau)d\tau + \int_0^t \frac{m}{q}(\tau)\, d\xi(\tau)\right) \tag{35}$$

where $\int$ denotes an Ito integral.

Hence

$$\frac{dP_{zx}}{dP_z}(\omega_2, \omega_1) = \frac{d\mu_{zx}}{d\mu_z}(z^t(\omega_2).\ X_\lambda(\omega_1)) = \frac{\mathscr{E}_x(\theta^t | X_\lambda = X_\lambda(\omega_1))}{\mathscr{E}_x(\theta^t)} \tag{36}$$

where

$$\theta^t - \exp\left(-\frac{1}{2}\int_0^t \frac{m^2}{q}(\tau)d\tau + \int_0^t \frac{m}{q}(\tau)\, dz(\tau, \omega_1)\right). \tag{37}$$

We note that $\frac{dP_{zx}}{dP_x}(\omega_2, \omega_1)$ is $\mathscr{A}_z \times \mathscr{A}_x$-measurable. Applying a general Bayes rule from ref. 16, we obtain

$$\frac{dP_{xz}}{dP_x}(\omega_1, \omega_2) = \frac{dP_{zx}}{dP_z}(\omega_2, \omega_1) \tag{38}$$

Let us denote the family of 2×2 orthogonal matrices by M_0. The set of induced Borel sets is denoted by $\mathscr{B}_0$. Let ν_{xz} be the conditional measure on $(M_0, \mathscr{B}_0)$ given $\mathscr{A}_z$, defined by $\nu_{xz}(A, \omega_2) = P(X_\lambda^{-1}(A) | \mathscr{A}_z^t)(\omega_2)$,

for $A \in \mathscr{B}_0$, $\omega_2 \in \Omega$. Let ν_x be the measure on $(M_0, \mathscr{B}_0)$ induced by X_λ. Then it is easily seen that

$$\frac{d\nu_{xz}}{d\nu_x}(X_\lambda(\omega_1),\ Z^t(\omega_2)) = \frac{dP_{xz}}{dP_x}(\omega_1, \omega_2) = \frac{\mathscr{E}_x(\theta^t | X_\lambda = X_\lambda(\omega_1))}{\mathscr{E}_x(\theta^t)} \tag{39}$$

where θ^t is defined by (37). Summarizing what has been shown, we have the following lemma.

Lemma 2: Consider the observation process described by (32). The conditional probability measure for the signal X_λ given the observation Z^t, ν_{xz}, is then absolutely continuous with respect to ν_x, the a priori measure for X_λ, and, for $Z^t \in C_2^t$ and $X \in M_0$,

$$\frac{d\nu_{xz}}{d\nu_x}(X,\ Z^t) = \frac{\mathscr{E}_x(\theta^t | X_\lambda = X)}{\mathscr{E}_x(\theta^t)} \tag{40}$$

where

$$\theta^t = \exp\left(-\frac{1}{2}\int_0^t \frac{m^2}{q}(\tau)d\tau + \int_0^t \frac{m}{q}(\tau)\,[Z'(\tau)dZ(\tau)]_{12}\right) \tag{41}$$

$$[Z'(\tau)dZ(\tau)]_{12} = [1,0]\,Z'(\tau)dZ(\tau)\begin{bmatrix}0\\1\end{bmatrix}. \tag{42}$$

If the density function of ν_x exists and is denoted by $p_{\underset{\sim}{X}_\lambda}(\cdot)$, then it follows from Lemma 1 that the density function $p_{\underset{\sim}{X}_\lambda}(\cdot | Z^t)$ of ν_{xz} exists and and can be expressed as follows:

$$p_{\underset{\sim}{X}_\lambda}(X | Z^t) = \frac{\mathscr{E}_x(\theta^t | X_\lambda = X)\, p_{\underset{\sim}{X}_\lambda}(X)}{\mathscr{E}_x(\theta^t)} \tag{43}$$

where θ^t is defined by (41). Let $x \in R^1$ be defined by $\exp Rx = X$ and $-\pi \le x < \pi$. Then by simple calculations,

$$p_{\underset{\approx}{x}_\lambda}(X|Z^t) = \frac{\mathscr{E}_x(\theta^t | x_\lambda = x \pm 2k\pi,\ k=1,2,\dots)\, p_{\underset{\approx}{x}_\lambda}(X)}{\mathscr{E}_x(\theta^t)}$$

$$= \sum_{k=-\infty}^{\infty} \frac{\mathscr{E}_x(\theta^t | x(\lambda) = x + 2k\pi)\, p_{x_\lambda}(x + 2k\pi)}{\mathscr{E}_x(\theta^t)} , \qquad (44)$$

where p_{x_λ} denotes the density function of $x(\lambda)$. This completes the proof of the following theorem.

<u>Theorem 4:</u> Consider the observation process described by (32). If the density function $p_{\underset{\approx}{x}_\lambda}$ of $X(\lambda)$ exists, then the conditional density function $p_{\underset{\approx}{x}_\lambda}(\cdot | Z^t)$ exists and can be expressed as follows

$$p_{\underset{\approx}{x}_\lambda}(X|Z^t) = \sum_{k=-\infty}^{\infty} p_{x_\lambda}(x+2k\pi | Z^t) = \sum_{k=-\infty}^{\infty} \frac{\mathscr{E}_x(\theta^t | x(\lambda) = x+2k\pi) p_{x_\lambda}(x+2k\pi)}{\mathscr{E}_x(\theta^t)} \qquad (45)$$

where θ^t is defined by (41), p_{x_λ} denotes the density function of $x(\lambda)$ and x is defined by $\exp Rx = X$ and $-\pi \le x < \pi$.

It is appropriate to remark that one can easily derive the stochastic partial differential equation for the conditional density $p_{\underset{\approx}{x}_\lambda}(X|Z^t)$ using Theorem 4 and the well-known equation (refs. 19, 20) for $p_{x_\lambda}(x+2k\pi | Z^t)$, $-\infty < k < \infty$. For economy of space, this equation will not be displayed. However we remark that when $m(x,t)$ is periodic in x with period 2π, the equation is in a form similar to the Stratonovich-Kushner equation with

p_{x_λ} replaced by $p_{\underset{\approx}{x}_\lambda}$.

Using Theorem 4 and the well-known fact (refs. 17, 18) that the smoothed and the predicted densities can be expressed explicitly in terms of filtering, we can easily obtain the following two corollaries.

Corollary 1: The conditional smoothed density $p_{\underset{\approx}{x}_\lambda}(X|Z^t)$, for $t_0 \leq \lambda \leq t$, may be expressed in terms of the conditional filtered density as follows:

$$p_{\underset{\approx}{x}_\lambda}(X|Z^t) = \sum_{k=-\infty}^{\infty} p_{x_\lambda}(x+2k\pi|Z^\lambda)\exp\left(\int_\lambda^t \frac{a_s}{q(s)}dI_s - \frac{1}{2}\int_\lambda^t \frac{a_s^2}{q(s)}ds\right) \tag{46}$$

where x is defined by $\exp Rx = X$ and $-\pi \leq x \leq \pi$ and

$$dI_s = [Z'(s)d\dot{Z}(s)]_{12} - \hat{m}(s)ds \tag{47}$$

$$a_s = \hat{m}(s|x_\lambda = x) - \hat{m}(s) \tag{48}$$

$$\hat{m}(s) = \mathscr{E}(m(s)|Z^s) \tag{49}$$

$$\hat{m}(s|x_\lambda = x) = \mathscr{E}(m(s)|Z^s,\ x_\lambda = x) \quad . \tag{50}$$

Corollary 2: Let X be a Markov process with given transition density $p_{\underset{\approx}{x}_\lambda}(X|x(t) = \xi)$. The conditional predicted density $p_{\underset{\approx}{x}_\lambda}(X|Z^t)$, for $t_0 \leq t \leq \lambda$ may be expressed in terms of the conditional filtered density as follows:

$$p_{\underset{\approx}{x}_\lambda}(X|Z^t) = \int_{-\infty}^{+\infty} p_{\underset{\approx}{x}_\lambda}(X|x(t) = \xi)\, p_{x_t}(\xi|Z^t)d\xi \quad . \tag{51}$$

3.3 Optimal Estimation

In the previous subsection, the conditional probability distributions were studied. A variety of estimation problems may be studied based on those conditional distributions. However, some estimation problems on the circle can be directly solved by using results in vector-space estimation theory. In this subsection, the well-established linear optimal estimation theory will be used to deduce recursive equations for optimal estimation on S^1 and thereby illustrate the approach.

The estimation problem which we will mainly be concerned with in this subsection is that of constructing a 2 x 2 orthogonal random matrix $\hat{X}(\lambda|t)$ as a $\mathscr{B}_1^t$-measurable functional of Z^t such that for a symmetric cost function ϕ defined by eq. (7), the following inequality holds for all $\mathscr{A}_z$-measurable 2 x 2 orthogonal random matrices M:

$$\mathscr{E}(\Phi(X(\lambda), \hat{X}(\lambda|t))|Z^t) \leq \mathscr{E}(\Phi(X(\lambda), M)|Z^t) \quad . \tag{52}$$

in which $\Phi(X_1, X_2) \triangleq \phi(\theta)$, θ being defined by $\exp R\theta = X_1^{-1}X_2$ and $-\pi \leq \theta < \pi$ (i.e. θ is the angle between X_1 and X_2).

We have seen, at the beginning of this section, that a continuous random process X on S^1 can be identified with a continuous random process x on R^1 via the bijective mapping $X = J(x)$. We now construct a signal process X on S^1 by injecting a linear diffusion x into S^1, x satisfying

$$dx(t) = a(t)x(t)\, dt + b^{1/2}(t)\, dv(t), \quad x(0) = 0 \tag{53}$$

where $b(t) > 0$, $\forall\, t \in T$, and v is a standard Brownian motion, independent of w, the observational noise. Applying the stochastic differentiation rule, we obtain the following stochastic differential equation for our signal

process $X = J(x)$:

$$dX(t) = -\frac{1}{2} b(t)X(t)dt + X(t)R\left\{a(t)\left[\int_0^t (\exp \int_s^t a(\tau)d\tau) b^{1/2}(s)dv(s)\right]dt + b^{1/2}(t)dv(t)\right\} \tag{54}$$

$$X(0) = I$$

where we note that $x(t) = \int_0^t (\exp \int_s^t a(\tau)d\tau) b^{1/2}(s)dv(s)$.

The observation process to be used in this subsection is taken to be Z, satisfying the stochastic differential equation:

$$dZ(t) = Z(t)\begin{bmatrix} -\frac{q(t)}{2} & c(t)x(t) \\ -c(t)x(t) & -\frac{q(t)}{2} \end{bmatrix} dt + Z(t)\begin{bmatrix} 0 & dw(t) \\ -dw(t) & 0 \end{bmatrix} \tag{55}$$

$$Z(0) = I$$

As shown in subsection 3.2., Z can be identified with $z = J^{-1}(Z)$ satisfying

$$dz(t) = c(t)x(t)dt + q^{1/2}(t)\, dw(t) \tag{56}$$

$$z(0) = 0$$

Note that the equations for X and Z are both bilinear in form. Moreover, z^t and Z^t generate the same σ-subfield $\mathcal{A}_z$ in $(\Omega, \mathcal{A}, P)$. Hence $\mathcal{E}(x(\lambda)|\mathcal{A}_z^t)$ is both a $\mathcal{B}_1^t$-measurable functional f_1 of z^t and a $\mathcal{B}_2^t$-measurable functional f_2 of Z^t, and

$$f_2(Z^t) = f_1(J^{-1}(Z^t)) \ . \tag{57}$$

Let $\hat{x}_{\lambda|t}$ and $\hat{x}(\lambda|t)$ denote $f_1(z^t) = \mathcal{E}(x(\lambda)|z^t)$ and $f_2(Z^t) = \mathcal{E}(x(\lambda)|Z^t)$ respectively.

We will first study the filtering problem, where $\sigma = t$. Then the Kalman-Bucy linear filtering theory yields immediately

$$d\hat{x}_{t|t} = a(t)\hat{x}_{t|t}dt + K(t)c(t)q^{-1}(t)\,(dz(t) - c(t)\hat{x}_{t|t}dt) \tag{58}$$

$$\hat{x}_{0|0} = 0$$

$$\dot{K}(t) = 2\,a(t)K(t) - c^2(t)q^{-1}(t)K^2(t) + b(t) \tag{59}$$

$$K(0) = 0$$

In view of (57), we obtain the following lemma, which not only leads to the solution of the above stated filtering problem but also applies directly to optimal frequency demodulation (see Section 5).

Lemma 3: Let the stochastic process (54) be the signal process and the stochastic process (55) be the observation process. Then the filtering equations are

$$d\hat{x}(t|t) = a(t)\hat{x}(t|t)dt + K(t)c(t)q^{-1}(t)\,([Z'(t)dZ(t)]_{12} - c(t)\hat{x}(t|t)dt)$$

$$\hat{x}(0|0) = 0 \tag{60}$$

$$\dot{K}(t) = 2\,a(t)K(t) - c^2(t)q^{-1}(t)K^2(t) + b(t) \tag{61}$$

$$K(0) = 0$$

and the conditional probability density is given by

$$p_{x_t}(x|Z^t) = \frac{1}{\sqrt{2\pi K(t)}}\,\exp\left[-\frac{1}{2K(t)}\,(x - \hat{x}(t|t))^2\right] \quad . \tag{62}$$

In view of Theorem 4, we see that $p_{\underset{\sim}{X}_t}(X|Z^t)$ is a folded normal density. By Theorem 2, it follows that $p_{\underset{\sim}{X}_t}(X|Z^t)$ is unimodal with mode at $\exp[\hat{x}(t|t)R]$ and is symmetric about it. We may now conclude from

Theorem 1 that for a cost function defined by (7),

$$\mathscr{E}(\Phi(X(t),\ \exp[\hat{x}(t|t)R])|Z^t) \leq \mathscr{E}(\Phi(X(t),\ M)|Z^t) \tag{63}$$

for any $\mathscr{A}_z$-measurable 2×2-dimensional orthogonal random matrix M. Since $\exp[\hat{x}(t|t)R]$ is easily seen to be a $\mathscr{B}_1^t$-measurable functional of Z^t, it follows that the optimal estimate of our signal process is

$$\hat{X}(t|t) = \exp[\hat{x}(t|t)R] \quad . \tag{64}$$

Differentiating this with respect to t yields

$$\begin{aligned} d\hat{X}(t|t) = & -\frac{1}{2}K^2(t)c^2(t)q^{-1}(t)\hat{X}(t|t)dt + \hat{X}(t|t)R((a(t) - K(t)c^2(t)q^{-1}(t)) \\ & \cdot \left[\int_0^t (\exp \int_s^t (a(\tau) - K(\tau)c^2(\tau)q^{-1}(\tau))d\tau)K(s)c(s)q^{-1}(s)[Z'(t)dZ(t)]_{12}\right]dt \\ & + K(t)c(t)q^{-1}(t)[Z'(t)dZ(t)]_{12}) \quad . \end{aligned} \tag{65}$$

Summarizing what has been shown, we obtain the following theorem.

<u>Theorem 5:</u> If the signal process X and the observation process Z on S^1 satisfy the following stochastic differential equations:

$$\begin{aligned} dX(t) = & -\frac{1}{2}b(t)X(t)dt + X(t)R(a(t)\left[\int_0^t (\exp \int_s^t a(\tau)d\tau)b^{1/2}(s)dv(s)\right]dt \\ & + b^{1/2}(t)dv(t)) \end{aligned} \tag{66}$$

$$X(0) = I$$

$$dZ(t) = Z(t)\begin{bmatrix} -\frac{q(t)}{2} & c(t)\left[\int_0^t X'(s)dX(s)\right]_{12} \\ -c(t)\left[\int_0^t X'(s)dX(s)\right]_{12} & -\frac{q(t)}{2} \end{bmatrix}dt$$

$$+ \ Z(t) \begin{bmatrix} 0 & dw(t) \\ -dw(t) & 0 \end{bmatrix} \tag{67}$$

$$Z(0) = I$$

where w and v are independent standard Brownian motions on R^1, then the optimal estimate $\hat{X}(t|t)$ in the sense of (52) satisfies the following stochastic differential equations:

$$d\hat{X}(t|t) = -\frac{1}{2} K^2(t)c^2(t)q^{-1}(t)\hat{X}(t|t)\,dt +$$

$$\hat{X}(t|t)R\Big((a(t) - K(t)c^2(t)q^{-1}(t))\Big[\int_0^t \Big(\exp \int_s^t (a(\tau) - K(\tau)c^2(t)q^{-1}(\tau))d\tau\Big)$$

$$\cdot K(s)c(s)q^{-1}(s)[Z'(s)dZ(s)]_{12}\Big]\,dt + K(t)c(t)q^{-1}(t)[Z'(t)dZ(t)]_{12}\Big) \tag{68}$$

$$\dot{K}(t) = 2a(t)K(t) - c^2(t)q^{-1}(t)K^2(t) + b(t) \tag{69}$$

$$K(0) = 0$$

The conditional probability density is given by

$$p_{\underset{\sim}{X}_t}(X|Z^t) = \frac{1}{\sqrt{2\pi K(t)}} \sum_{k=-\infty}^{\infty} \exp\left[-\frac{1}{2K(t)}(x + 2k\pi - \hat{x}(t|t))^2\right]$$

$$= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \exp\left[-\frac{k^2K^2(t)}{2}\right] \cos k(x - \hat{x}(t|t)) \tag{70}$$

where x is defined by $\exp R\,x = X$ and $-\pi \leq \quad \pi$.

The expected error $\mathcal{E}(\Phi(X(t), \hat{X}(t|t)))$ of the optimal estimate $\hat{X}(t|t)$ can be obtained by straightforward computation with the aid of (70). Some

examples are given in the following corollary.

Corollary: Let θ be defined by $\exp R\theta = X$ and $-\pi \leq \theta < \pi$. Then

(i) for $\phi(\theta) = 1 - \cos\theta$,

$$\mathscr{E}(\Phi(X(t), \hat{X}(t|t))) = 1 - \exp(-\tfrac{1}{2}K(t)) \tag{71}$$

(ii) for $\phi(\theta) = (1 - \cos\theta)^2$,

$$\mathscr{E}(\Phi(X(t), \hat{X}(t|t))) = \frac{3}{2} - 2\exp(-\frac{K(t)}{2}) + \frac{1}{2}\exp(-2K(t)) \tag{72}$$

(iii) for $\phi(\theta) = \rho(\theta)$, the Riemannian metric,

$$\mathscr{E}(\Phi(X(t), \hat{X}(t|t))) = \frac{1}{2} - \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{1}{(2k+1)^2}\exp\left[-\frac{(2k+1)^2K(t)}{2}\right] \tag{73}$$

(iv) for $\phi(\theta) = \rho^2(\theta)$,

$$\mathscr{E}(\Phi(X(t), \hat{X}(t|t))) = \frac{2\pi^2}{3} - 4\left[\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k^2}\exp\left(-\frac{k^2K(t)}{2}\right)\right]. \tag{74}$$

We recall that $K(t) = \mathscr{E}(x(t) - \hat{x}_{t|t})^2$. From this Corollary, it can be seen that for the cases (i) $\sim$ (iv), $\mathscr{E}(\Phi(X(t), \hat{X}(t|t)))$ is a monotone increasing function of $\mathscr{E}(x(t) - \hat{x}_{t|t})^2$. It has been shown in Section 2 that this property holds for all ϕ defined by (7).

We note that the optimal filtering equations (68) and (69) are complex in form. The concept of the filtering procedure, however, is quite simple, and is best illustrated by the block diagram of Fig. 4.

The observation process dZ first goes through a nonlinear transformer. The transformed process $[Z'dZ]_{12}$ then goes through a Kalman-Bucy linear filter. Then we inject the filtered process $\hat{x}(t|t)$ into S^1 via the

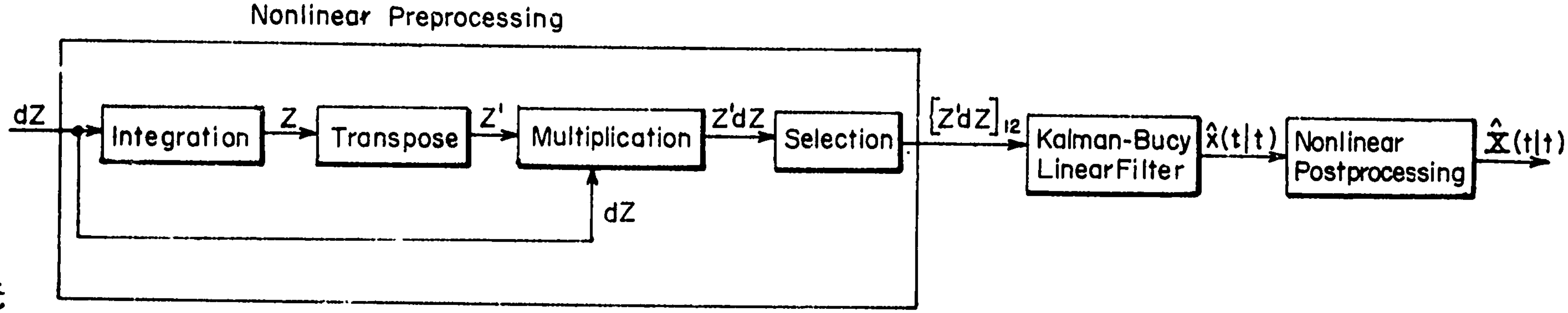

Figure 4 Block Diagram for Optimal Filtering

injection mapping J. The output $\hat{X}(t|t)$ of the nonlinear injector is the desired estimate.

The same approach can be used to solve the smoothing and prediction problems. The solution to the prediction problem is trivial and hence omitted here. For the smoothing problem, we first recall (ref. 22) that for $0 \leq \lambda \leq t$,

$$\hat{x}_{\lambda|t} = \hat{x}_{\lambda|\lambda} + K(\lambda) \int_{\lambda}^{t} (\exp \int_{\lambda}^{s} (a(\tau) - K(\tau)c^{2}(\tau)q^{-1}(\tau))d\tau)c(s)q^{-1}(s)(dz(s) - c(s)\hat{x}_{s|s}ds) \quad . \tag{75}$$

By (57), it follows that

$$\hat{x}(\lambda|t) = \hat{x}(\lambda|\lambda) + K(\lambda) \int_{\lambda}^{t} (\exp \int_{\lambda}^{s} (a(\tau) - K(\tau)c^{2}(\tau)q^{-1}(\tau))d\tau \cdot c(s)q^{-1}(s)([Z'(s)dZ(s)]_{12} - c(s)\hat{x}(s|s)ds) \quad . \tag{76}$$

We note that the conditional probability distribution of $x(\lambda)$ given Z^t is Gaussian. From Theorem 4, it follows that $p_{\underset{\sim}{X}_\lambda}(X|Z^t)$ is a folded-Gaussian density and hence unimodal. As in the filtering case,

$$\hat{X}(\lambda|t) = \exp(\hat{x}(\lambda|t)R) \quad . \tag{77}$$

Substituting (76) into (77) thus yields

$$\hat{X}(\lambda|t) = \hat{X}(\lambda|\lambda) \exp \left\{ RK(\lambda) \int_{\lambda}^{t} (\exp \int_{\lambda}^{s} (a(\tau) - K(\tau)c^{2}(\tau)q^{-1}(\tau)d\tau) \cdot c(s)q^{-1}(s) \left([Z'(s)dZ(s)]_{12} - c(s) \int_{0}^{s} [\hat{X}'(\tau|\tau)d\hat{X}(\tau|\tau)]_{12}\, ds \right) \right\} \tag{78}$$

where we have used the identity $\hat{x}(s|s) = \int_0^s [\hat{X}'(\tau|\tau)dX(\tau|\tau)]_{12}$. Summarizing what has been shown, we obtain the following theorem.

Theorem 6: If the signal process and the observation process are the same as in Theorem 5, then the optimal estimate, $\hat{X}(\lambda|t)$, $0 \le \lambda \le t$, in the sense of (52), is given by

$$\hat{X}(\lambda|t) = \hat{X}(\lambda|\lambda) \exp\left\{RK(\lambda) \int_\lambda^t \left(\exp \int_\lambda^s (a(\tau) - K(\tau)c^2(\tau)q^{-1}(\tau))d\tau\right)\right.$$

$$\left. \cdot\, c(s)q^{-1}(s)\left([Z'(s)dZ(s)]_{12} - c(s) \int_0^s [\hat{X}'(\tau|\tau)d\hat{X}(\tau|\tau)]_{12}\, ds\right)\right\}, \tag{79}$$

where $\hat{X}(\tau|\tau)$, $K(\tau)$ can be obtained from (68) and (69).

The conditional probability density of $X(\lambda)$ given Z^t, the expected errors $\mathscr{E}(\Phi(X(\lambda), \hat{X}(\lambda|t))$, the stochastic equations for $\hat{X}(\lambda|t)$ for fixed-point smoothing, fixed-lag smoothing, and fixed interval smoothing can all be easily obtained by straightforward computations. They are left to the interested readers.

3.4 Random Initial State

In the previous subsections, the initial state of the signal process X is assumed to be $X(0) = I$, the identity matrix. This is obviously not a practical assumption in some applications. In this subsection we will consider the case in which the initial state is a random variable. We will denote the signal process by Y in this subsection, and assume that $Y(0) = Y_0$ is a random variable independent of the observational noise w.

We observe that the input to the observation process (32) at time t is not the dynamical state of the signal. It is the angle that the rotational

process represented by the signal has swept over the time interval $[0, t]$. Taking this viewpoint, our present problem can be solved with some modification to the previous results.

Let $y(t)$ denote the angle that the signal Y has swept over $[0, t]$. It is easily seen that

$$y(t) = \left[\int_0^t Y'(s)dY(s)\right]_{12} \quad . \tag{80}$$

Define a rotational process X by

$$X(t) = Y_0^{-1}Y(t) \quad . \tag{81}$$

Then $X(0) = I$ and, as before, we may define

$$x(t) = (J^{-1}(X))(t) = \left[\int_0^t X'(s)dX(s)\right]_{12} \quad . \tag{82}$$

We note that $x(t) = y(t)$. In other words, the angles swept by X and by Y over $[0, t]$ are the same. Hence (32) can also be used as the observation process for our present problem. The conditional distribution of $X(\lambda)$ given observation Z^t of the form given in (32) can be determined by the application of the previous results.

We note that Y_0 and $X(\lambda)$ are conditionally independent given Z^t. If the distribution of Y_0 and the conditional distribution of $X(\lambda)$ given Z^t are both folded normal, then the following lemma easily leads to the conclusion that the optimal estimate $\hat{Y}(\lambda|t)$ of $Y(\lambda)$ given Z^t is equal to $\hat{Y}_0\hat{X}(\lambda|t)$, where $\hat{Y}_0$ is the mode of the distribution of Y_0 and $\hat{X}(\lambda|t)$ is the mode of the conditional distribution of $X(\lambda)$ given Z^t.

<u>Lemma 4:</u> Let A and B be two independent 2×2 orthogonal random

matrices which have folded normal distributions with modes $\hat{A}$ and $\hat{B}$ respectively. Then AB is a 2 x 2 orthogonal random matrix which has a folded normal distribution with mode equal to $\hat{A}\hat{B}$.

Proof. It is easily seen that there exist unique real-valued normal random variables a and b such that $\mathscr{E}a, \mathscr{E}b \in [-\pi, \pi)$, $A = \exp Ra$, and $B = \exp Rb$. Then $AB = \exp R(a+b)$. Obviously a+b is a normal random variable. Hence AB is folded normal and the mode of AB is $\exp[R\mathscr{E}(a+b)] = \exp[R\mathscr{E}(a)] \cdot \exp[R\mathscr{E}(b)] = \hat{A}\hat{B}$. ∎

3.5 Multichannel Estimation

The results of the previous subsections can be extended to a large class of problems -- those involving processes evolving on abelian Lie groups. It is well known (ref. 23) that a given abelian Lie group G is isomorphic to the direct product of a number of copies of the circle and a number of copies of the real line, i. e.

$$G \approx R^n x (S^1)^m$$

where $(S^1)^m$ is usually called a "torus". The diffusion processes on this type of space have been used to model some interesting satellite and pendulum systems in ref. 46. Analogous to (28), a bijective mapping $J_{nm}:(C_1^s)^{n+m} \longrightarrow (C_1^s)^n \times (C_2^s)^m$ is defined by

$$(J_{nm}(a))(t) = [a_1(t), \ldots, a_n(t), \ (J(a_{n+1}))(t), \ldots, (J(a_{n+m}))(t)] \tag{83}$$

for $a \in (C_1^s)^{n+m}$, a_i being the ith component of a. Thus a continuous random signal process on G which is described by an $\mathscr{A}$-measurable function $X:\Omega \longrightarrow (C_1^s)^n \times (C_2^s)^m$ corresponds to a unique continuous random signal process on R^{n+m} which is described by an $\mathscr{A}$-measurable function

$x: \Omega \to (C_1^s)^{m+n}$ such that

$$X(t) = (J_{nm}(x))(t), \qquad t \in [0, S]. \qquad (84)$$

The mathematical model for the sensor can be obtained by first using J_{nm} to inject the following vector random differential equation into $R^n \times (S^1)^m$

$$\begin{aligned} dz(t) &= m(x(t), t)dt + dv(t) \\ z(0) &= 0 \end{aligned} \qquad , \qquad (85)$$

and then differentiating $Z(t) = (J_{nm}(z))(t)$ by the stochastic differentiation rule to obtain a set of stochastic differential equations of which the first n equations are the same as the first n equations of (85) and the last m equations are bilinear 2×2 matrix differential equations in the form of (32). This calculation is straightforward and thus we will not display those sensor equations. Because of the bijective property of J_{nm}, it is clear that the estimation analysis in the previous subsections can be easily generalized to this general abelian case with little modification. For the special case in which x is a linear diffusion and $m(x(t), t)$ is a linear function of $x(t)$, what has been shown simply asserts that the domain of the celebrated Kalman-Bucy filter includes estimation on abelian Lie groups.

3.6 Examples

To illustrate the ideas of the preceding discussions, we present the following examples.

Example 1: Consider a cylindrical shaft of unit radius being spun about its longitudinal axis by an electric motor. We assume that the total rotation of the shaft, x_1, is related to the driving force u by the differential

equation

$$\ddot{x}_1 + \dot{x}_1 + x_1 = u \qquad ,$$

with both $x_1(0)$ and $\dot{x}_1(0)$ equal to zero. The last term on the left-hand side of this equation can be thought of as a torsional spring effect, which helps to stabilize the servo loop that drives the shaft. The driving force u consists of a known driving force and a disturbance. The known driving force adds neither difficulty to the analysis nor complexity to the solution. Thus, for simplicity, we assume that the known driving force is zero and that the disturbance is white Gaussian noise -- i.e. $u = \dot{v}$, where v is a standard one-dimensional Brownian motion. Setting $x_2 = \dot{x}_1$ we obtain the vector stochastic differential equation

$$dx(t) = Ax(t)dt + Bdv(t) \qquad x(0) = 0 \quad ,$$

where

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Suppose we wish to estimate the orientation of the shaft. The orientation is determined by the quantities $\sin x_1(t) = \sin \int_0^t x_2(\tau)d\tau$ and $\cos x_1(t) = \cos \int_0^t x_2(\tau)d\tau$. Suppose also that we have some means of measuring these quantities, but that noise corrupts the measurements, so that our actual measurements are $z_1(t) \triangleq \cos(\int_0^t x_2(\tau)d\tau + w(t))$ and $z_2(t) \triangleq \sin(\int_0^t x_2(\tau)d\tau + w(t))$ where w is a standard Brownian motion

process independent of v. Using the Ito differential rule, we obtain the sensor equations

$$dz_1(t) = -\frac{1}{2} z_1(t)dt - x_2(t)z_2(t)dt - z_2(t)dw(t); \; z_1(0) = 1$$

$$dz_2(t) = -\frac{1}{2} z_2(t)dt + x_2(t)z_1(t)dt + z_1(t)dw(t); \; z_2(0) \quad 0$$

Using the results of this section, we have the following cptimal filtering equations

$$d\hat{x}(t|t) = A\hat{x}(t|t)dt + K(t)c'[y_1(t)dy_2(t) - y_2(t)dy_1(t) - c\hat{x}(t|t)dt]$$

$$\hat{x}(0|0) = 0$$

where

$$c = [0,1]$$

and K is the 2 x 2 solution of

$$\dot{K}(t) = AK(t) + K(t)A' - K(t)c'cK(t) + B \; B'; K(0) = 0$$

Finally, the optimal estimate of the orientation -- i. e. the optimal estimate of

$$X_1(t) = \exp(x_1(t)R) = \begin{bmatrix} \cos x_1(t) & \sin x_1(t) \\ -\sin x_1(t) & \cos x_1(t) \end{bmatrix}$$

is

$$\hat{X}_1(t|t) = \exp(\hat{x}_1(t|t)R)$$

The steady state filter has the same form as the time-varying filter, but K(t) is replaced by the positive definite solution, K_∞, of the algebraic Riccati equation

$$AK_{\infty} + K_{\infty}A' - K_{\infty}c'cK_{\infty} + BB' = 0$$

The solution to this is

$$K_{\infty} = \begin{bmatrix} \sqrt{2} - 1 & 0 \\ 0 & \sqrt{2} - 1 \end{bmatrix}$$

If we formally divide dz_1 and dz_2 by dt and take $\dot{z}_1$ and $\dot{z}_2$ to be our measurements, we get the following block diagram (Figure 5) for the signal process, observation process, nonlinear preprocessor, optimal filter, and nonlinear postprocessor.

Example 2: In this example, the nonlinear signal process and the nonlinear observation process of a certain system turn out to be processes taking values on the abelian Lie groups $S^1 \times R^2$ and $S^1 \times R^1$, respectively. The signal process is four-dimensional, satisfying

$$dx_1 = -\frac{1}{2} x_1 dt - x_2 x_3 dt - x_2 dv$$

$$dx_2 = -\frac{1}{2} x_2 dt + x_1 x_3 dt + x_1 dv$$

$$\dot{x}_3 = -\int_0^t x_3(s)ds + v$$

$$\dot{x}_4 \quad x_3 + x_4$$

$$x_1(0) = 1, \quad x_2(0) = x_3(0) = x_4(0) = 0 \ .$$

The sensor equations are

$$dz_1 = -\frac{1}{2} z_1 dt - (2x_4 + x_3 - \int_0^t x_1(s)dx_2(s) + \int_0^t x_2(s)dx_1(s))\, z_2 dt - z_2 dw_1$$

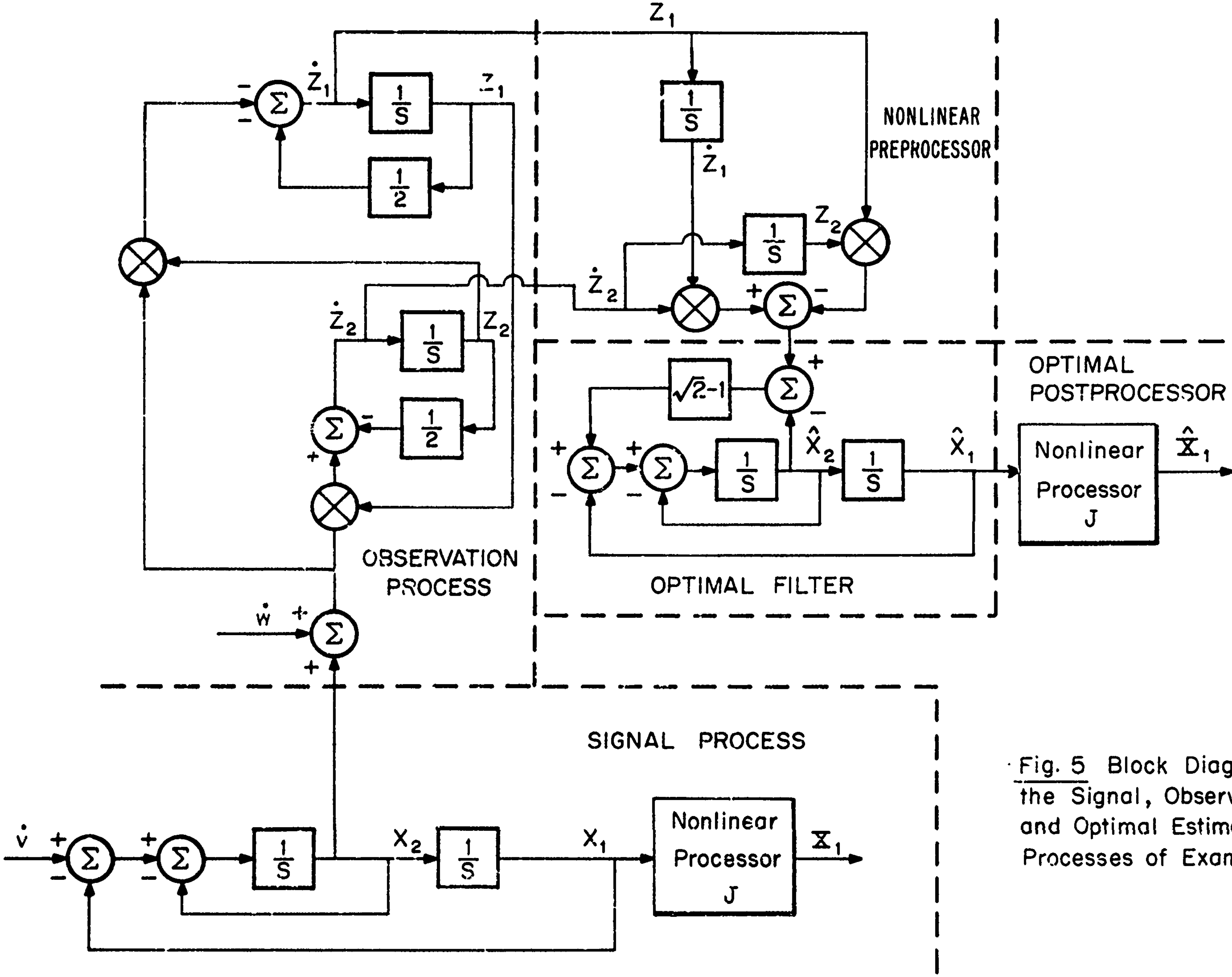

Fig. 5 Block Diagram of the Signal, Observation, and Optimal Estimation Processes of Example 1.

$$dz_2 = -\frac{1}{2} z_2 dt + \left(2x_4 + x_3 - \int_0^t x_1(s)dx_2(s) + \int_0^t x_2(s)dx_1(s)\right)$$

$$\cdot\ z_1 dt + z_1 dw_1$$

$$dz_3 = \left(\int_0^t x_1(s)dx_2(s) - \int_0^t x_2(s)dx_1(s)\right) dt + dw_2$$

$$z_1(0) = 1, \qquad z_2(0) = z_3(0) = 0 \qquad ,$$

where w_1 and w_2 are standard Brownian motions independent of each other and of v. Our problem is to find the least-squares estimate $\hat{x}$ under the constraint $\hat{x}_1^2 + \hat{x}_2^2 = 1$. Rearrangements of the first two signal equations yield

$$\begin{bmatrix} dx_1 & dx_2 \\ -dx_2 & dx_1 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix} dt + \begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix} R(x_3 dt + dv) \quad .$$

Comparing this equation with (54), we see that its solution describes a rotational process with a single degree of freedom. Let y(t) denote the total rotation completed at t. Then $x_1(t) = \cos y(t)$, $x_2(t) = \sin y(t)$, $dy = x_3 dt + dv - x_1 dx_2 - x_2 dx_1$, and the first two sensor equations become, after some rearrangements,

$$\begin{bmatrix} dz_1 & dz_2 \\ -dz_2 & dz_1 \end{bmatrix} = \begin{bmatrix} z_1 & z_2 \\ -z_2 & z_1 \end{bmatrix} \left(-\frac{1}{2} I dt + R(2x_4 + x_3 - y)\, dt + R\, dw_1\right)$$

We note that the system is not observable with just the S^1 observation pair $\{z_1, z_2\}$ or with just the R^1 observation z_3, but that the system is observable when both observation processes are present.

Following the approach developed in this section, we first obtain the following optimal filtering equations:

$$\begin{bmatrix} d\hat{y} \\ d\hat{x}_3 \\ d\hat{x}_4 \end{bmatrix} = A \begin{bmatrix} \hat{y} \\ \hat{x}_3 \\ \hat{x}_4 \end{bmatrix} + KC' \left\{ \begin{bmatrix} z_1 dz_2 - z_2 dz_1 \\ dz_3 \end{bmatrix} - C \begin{bmatrix} \hat{y} \\ \hat{x}_3 \\ \hat{x}_4 \end{bmatrix} dt \right\}$$

$$[\hat{y}(0), \ \hat{x}_3(0), \ \hat{x}_4(0)] = 0$$

where

$$\dot{K} = AK + KA' - KC'CK + BB'$$

$$K(0) = 0$$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix} .$$

With help of previous results, we see that

$$\mathscr{E}(1 - \cos(y - \hat{y})) \leq \mathscr{E}(1 - \cos(y - \xi))$$

for all z^t-measurable ξ. Hence

$$\begin{aligned} \frac{1}{2} \mathscr{E}[(\cos y - \cos \hat{y})^2 + (\sin y - \sin \hat{y})^2] \\ = 1 - \mathscr{E}[\cos(y - \hat{y})] \leq 1 - \mathscr{E}[\cos(y - \xi)] \\ = \frac{1}{2} \mathscr{E}[(\cos y - \cos \xi)^2 + (\sin y - \sin \xi)^2] \end{aligned}$$

for all z^t-measurable ξ. This shows that the least-squares estimates $\hat{x}_1$ and $\hat{x}_2$ under the constraint $\hat{x}_1^2 + \hat{x}_2^2 = 1$ are given by

$$\hat{x}_1 = \cos \hat{y}$$
$$\hat{x}_2 = \sin \hat{y}$$

The block diagram of the optimal filter is given in Figure 6.

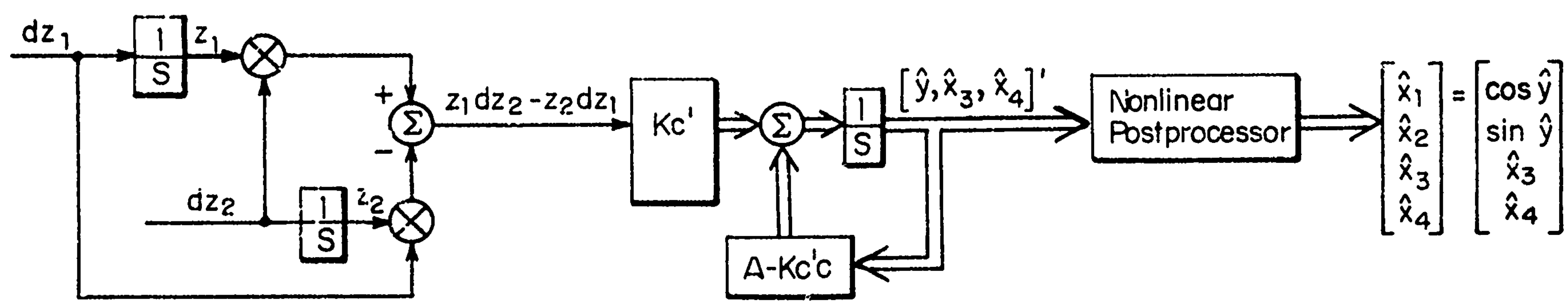

Fig. 6. Block diagram of the optimal filter for Example 2

4. Discrete Time Estimation

We now wish to examine the problem of estimating a random process on S^1, given a series of discrete measurements. A natural model for the measurement process is a discrete approximation to the continuous measurement process discussed in the preceding section. We approximate the continuous measurement process

$$dz(t) = m(x(t), t)\, dt + \sqrt{q(t)}\, dw(t)$$

$$Z(t) = (J(z))(t)$$

by the discrete equations

$$\Delta y_k = y_k - y_{k-1} = m_k(x_k)\Delta t + \sqrt{q_k}\, \Delta w_k$$

$$Y_k = \exp(y_k R)$$

where Δt is the inter-measurement time, $x_k = x(k\Delta t)$, $q_k = q(k\Delta t)$, $m_k(\cdot) = m(\cdot, k\Delta t)$, and $\Delta w_k = w(k\Delta t) - w((k-1)\Delta t)$.

We can rewrite the Y_k equation as

$$Y_k = Y_{k-1} \exp(\Delta y_k R) , \tag{86}$$

and we see that, given the measurements $Y_1, \ldots, Y_{k-1}$ the new information contained in Y_k is equivalent to the new information in $Y_{k-1}^{-1} Y_k$. This information is easily seen to be equivalent to the knowledge of

$$\tilde{\Delta} y_k \triangleq \Delta y_k \bmod 2\pi , \tag{87}$$

where we adopt the convention $\tilde{\Delta} y_k \in [-\pi, \pi)$.

It is here that we see a marked difference between the discrete and continuous problems. In the continuous time problem, the continuity of the stochastic processes results in our knowing $dy(t)$, not just $dy(t) \bmod 2\pi$. However, in the discrete problem, the ambiguity associated with our lack of knowledge of the number of rotations that occur in the Δt between measurements, is reflected in the fact that our information is just $\Delta y_k \bmod 2\pi$.

With this discretization as motivation, in subsection 4.1 we will formulate a class of single stage estimation problems on S^1, and will derive conditional density equations that lend themselves to a relatively simple physical interpretation when considered alongside the preceding comments. In addition, extensions to the multistage problem are discussed.

The results of this subsection provide a striking example of a class of systems for which the continuous time problem is decidedly <u>less</u> complex than the discrete time problem. Thus practical suboptimal schemes are necessary in the discrete time case. To this end, an appendix has been included, in which the relationship between the discrete and continuous problems is discussed. Motivated by this discussion, several suboptimal schemes for the discrete problem are discussed at the end of subsection 4.1.

In subsection 4.2, we will use Fourier series analysis to study a more general discrete time estimation problem on the circle. The form of the conditional density equations will suggest a simple method for designing suboptimal filters for any estimation problem on S^1.

4.1 Conditional Distributions on S^1 and Optimal Estimation

Suppose we are given a random variable x, taking values in R^1, with a priori density $p_x(\alpha)$. We can "project" this variable onto the circle by the equation

$$\theta = x \bmod 2\pi .$$

The a priori density (with respect to the standard (Haar) measure on S^1) for θ is given by the associated projection map

$$p_\theta(\tilde{\alpha}) = \sum_{n=-\infty}^{+\infty} p_x(\tilde{\alpha} + 2n\pi) \quad ; \quad \tilde{\alpha} \in [-\pi, \pi)$$

We suppose that a measurement of the form

$$\tilde{y} = (m(x) + v) \bmod 2\pi \quad ; \quad \tilde{y} \in [-\pi, \pi)$$

is taken, where v is a random variable on R^1, independent of x (and thus θ), with density $p_v(\nu)$, and $m: R^1 \to R^1$ is a Borel measurable function. We also define the auxiliary, unobtainable "measurement"

$$y = m(x) + v$$

which has a density function given by

$$p_y(\beta) = \int_{-\infty}^{+\infty} p_v(\beta - m(u))\, p_x(u)\, du .$$

Then we have the density for $\tilde{y}$:

$$p_{\tilde{y}}(\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} p_y(\tilde{\beta} + 2n\pi) \quad ; \quad \tilde{\beta} \in [-\pi, \pi)$$

We wish to compute the conditional density $p_{\theta|\tilde{y}}(\tilde{\alpha}|\tilde{\beta})$, or, equivalently, the density $p_{x|\tilde{y}}(\alpha|\tilde{\beta})$. Also, we wish to know if this density has a particularly nice form if m is linear and x and v are normally distributed.

We will derive somewhat more general results, and will apply them to this problem. The arguments in this section are measure-theoretic in nature, and are summarized in the statements of Theorem 7, Theorem 8, and their corollaries. The solution to the specific problem stated above is given in the statements of the two corollaries to Theorem 8.

We consider the probability space $(R^1, \mathscr{A}^1, P_y)$ where $\mathscr{A}^1$ is the σ-algebra of Borel measurable subsets of R^1, and P_y is any probability measure on $\mathscr{A}^1$. We define two random variables on this space.

$$y(\omega) = \omega$$

$$\tilde{y}(\omega) = \omega \bmod 2\pi \qquad (\tilde{y} \in [-\pi, \pi))$$

Then $\mathscr{A}_y$, the σ-field generated by y, is $\mathscr{A}$, and $\mathscr{A}_{\tilde{y}}$ (defined analogously) consists of the following sets

$$\tilde{A} \in \mathscr{A}_{\tilde{y}} \Leftrightarrow \tilde{A} \cap [-\pi, \pi) \triangleq A \in \mathscr{A} \quad \text{and}$$

$$\tilde{A} = \bigcup_{n=-\infty}^{+\infty} (A + 2n\pi)$$

i.e. $\tilde{A}$ is "periodic" in form and thus is determined by A.

We define a sequence of measures on the measurable space $([-\pi,\pi), \mathscr{S})$, where $\mathscr{S}$ is the σ-field of Borel subsets of $[-\pi,\pi)$:

$$P_y^n(S) \triangleq P_y(S + 2n\pi)$$

$$\tilde{P}_y(S) \triangleq \sum_{n=-\infty}^{+\infty} P_y^n(S) \qquad S \in \mathscr{S}$$

clearly $\tilde{P}_y(S) = 0 \Rightarrow P_y^n(S) = 0$, which means P_y^n is absolutely continuous with respect to $\tilde{P}_y$ $(P_y^n << \tilde{P}_y)$, and thus, by the Radon-Nikodym Theorem, [24], [25], for each n there exists an $\mathscr{S}$-measurable function $dP_y^n/d\tilde{P}_y$, such that

$$P_y^n(S) = \int_S \frac{dP_y^n}{d\tilde{P}_y}(\omega)\, d\tilde{P}_y(\omega) \qquad \forall S \in \mathscr{S}$$

We wish to compute the conditional probability measure $P_{y|\tilde{y}}$ for y given $\tilde{y}$. The following theorem shows that this conditional measure can be expressed in a form that reflects our uncertainty as to the number of integral multiples of 2π that separate the values of y and $\tilde{y}$.

Theorem 7: The conditional probability measure

$$P_{y|\tilde{y}}(C|\tilde{\beta}) = P(y \in C|\tilde{y} = \tilde{\beta}) \quad ; \quad \tilde{\beta} \in [-\pi,\pi), \quad C \in \mathscr{A}$$

can be expressed in the form

$$P_{y|\tilde{y}}(C|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} \chi_C(\tilde{\beta} + 2n\pi)\frac{dP_y^n}{d\tilde{P}_y}(\tilde{\beta}) \tag{88}$$

where χ_C is the characteristic function for C.

Proof: From the definition of conditional expectation, [26], [27], [28], we have that $P_{y|\tilde{y}}(C|\beta)$ is the unique (up to with-probability-one equivalence) $\mathscr{A}_{\tilde{y}}$-measurable function, such that for any $\tilde{A} \in \mathscr{A}_{\tilde{y}}$

$$\int_{\tilde{A}} \chi_C(\omega)\, dP_y(\omega) = \int_{\tilde{A}} P_{y|\tilde{y}}(C|\tilde{y}(\omega))\, dP_y(\omega) \quad . \tag{89}$$

The left-hand side of (89) equals

$$\sum_{n=-\infty}^{+\infty} \int_{A+2n\pi} \chi_C(\omega)\, dP_y(\omega) = \sum_{n=-\infty}^{+\infty} \int_A \chi_C(\xi + 2n\pi)\, dP_y^n(\xi)$$

where $A = \tilde{A} \cap [-\pi, \pi)$. This last equality follows from the definition of P_y^n and the obvious relationships among $\mathscr{A}$- and $\mathscr{G}$-measurability.

Now $P_y^n << \tilde{P}_y$, so

$$\sum_{n=-\infty}^{+\infty} \int_A \chi_C(\xi + 2n\pi)\, dP_y^n(\xi) =$$

$$= \sum_{n=-\infty}^{+\infty} \int_A \chi_C(\xi + 2n\pi) \frac{dP_y^n}{d\tilde{P}_y}(\xi)\, d\tilde{P}_y(\xi)$$

$$= \int_A \left[\sum_{n=-\infty}^{+\infty} \chi_C(\xi + 2n\pi) \frac{dP_y^n}{d\tilde{P}_y}(\xi) \right] d\tilde{P}_y(\xi) \quad , \tag{90}$$

where we have used the Radon-Nikodym Theorem, the fact that $dP_y^n/d\tilde{P}_y \geq 0$ a.e. $(\tilde{P}_y)$, and the monotone convergence theorem, [24].

Similarly, the right-hand side of (89) is equal to

$$\sum_{n=-\infty}^{+\infty} \int_{A+2n\pi} P_{y|\tilde{y}} (C|\tilde{y}(\omega)) \, dP_y(\omega)$$

$$= \sum_{n=-\infty}^{+\infty} \int_A P_{y|\tilde{y}} (C|\tilde{y}(\xi)) \, dP_y^n (\xi)$$

where we have used the periodicity of $\tilde{y}(\xi)$. Again using the Radon-Nikodym and monotone convergence theorems, this last expression is equal to

$$\sum_{n=-\infty}^{+\infty} \int_A P_{y|\tilde{y}} (C|\tilde{y}(\xi)) \frac{dP_y^n}{d\tilde{P}_y} (\xi) \, d\tilde{P}_y (\xi)$$

$$= \int_A P_{y|\tilde{y}} (C|\tilde{y}(\xi)) \left[\sum_{n=-\infty}^{+\infty} \frac{dP_y^n}{d\tilde{P}_y} (\xi) \right] d\tilde{P}_y (\xi)$$

Clearly $\tilde{P}_y$ is a probability measure on $[-\pi, \pi)$, since

$$\tilde{P}_y ([-\pi,\pi)) = \sum_{n=-\infty}^{+\infty} P_y^n ([-\pi,\pi))$$

$$= \sum_{n=-\infty}^{+\infty} P_y ([-\pi,\pi) + 2n\pi) = P_y((-\infty, \infty)) = 1$$

Also, on any $\mathscr{P}$-measurable set S ,

$$\tilde{P}_y(S) = \sum_{n=-\infty}^{+\infty} P_y^n(S) = \sum_{n=-\infty}^{+\infty} \int_S \frac{dP_y^n}{d\tilde{P}_y}(\xi)\, d\tilde{P}_y(\xi)$$

$$= \int_S \left[\sum_{n=-\infty}^{+\infty} \frac{dP_y^n}{d\tilde{P}_y}(\xi) \right] d\tilde{P}_y(\xi) \ ,$$

and, since $\tilde{P}_y$ is a finite measure, this implies that

$$\sum_{n=-\infty}^{+\infty} \frac{dP_y^n}{d\tilde{P}_y}(\xi) = 1 \qquad \text{a. e.} \quad (\tilde{P}_y)$$

Thus, the right-hand side of (89) is

$$\int_A P_{y|\tilde{y}}(C|\tilde{y}(\xi))\, d\tilde{P}_y(\xi) \tag{91}$$

Comparing (90) and (91) we see that the conditional measure is given by

$$P_{y|\tilde{y}}(C|\tilde{y}(\xi)) = \sum_{n=-\infty}^{+\infty} \chi_C(\xi + 2n\pi) \frac{dP_y^n}{d\tilde{P}_y}(\xi) \ .$$

But this is defined for $\xi \in [-\pi, \pi)$, and in this case $\tilde{y}(\xi) = \xi$. Thus

$$P_{y|\tilde{y}}(C|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} \chi_C(\tilde{\beta} + 2n\pi) \frac{dP_y^n}{d\tilde{P}_y}(\tilde{\beta}) \qquad \blacksquare$$

We note that for fixed $\tilde{\beta}$, $P_{y|\tilde{y}}(C|\tilde{\beta})$ is a sum of Dirac measures concentrated at the points $\beta + 2n\pi$, where

$$P(y = \tilde{\beta} + 2n\pi | \tilde{y} = \tilde{\beta}) = \frac{dP_y^n}{d\tilde{P}_y}(\tilde{\beta})$$

Thus, in terms of δ-functions, we can write the conditional "density"

$$p_{y|\tilde{y}}(\xi|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} \frac{dP_y^n}{d\tilde{P}_y}(\tilde{\beta})\ \delta(\xi - \tilde{\beta} - 2n\pi) \qquad (92)$$

Corollary: Suppose P_y is absolutely continuous with respect to Lebesque measure λ :

$$P_y(A) = \int_A p_y(\xi)\, d\lambda(\xi)$$

Then the conditional "density" is given by

$$p_{y|\tilde{y}}(\xi|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} \frac{p_y(\tilde{\beta}+2n\pi)}{\sum_{n=-\infty}^{+\infty} p_y(\tilde{\beta}+2k\pi)}\ \delta(\xi - \tilde{\beta} - 2n\pi) \qquad (93)$$

$$= \frac{p_y(\xi)}{\sum_{k=-\infty}^{+\infty} p(\xi+2k\pi)}\ \delta(\tilde{\beta} - (\xi \bmod 2\pi)) \qquad (94)$$

Proof: It is easy to see that P_y^n is absolutely continuous with respect to Lebesque measure (also called λ) on $[-\pi,\pi)$, and

$$\frac{dP_y^n}{d\lambda}(\eta) = p_y(\eta+2n\pi)$$

is a version of the Radon-Nikodym derivative. Clearly $p_y \geq 0$ a. e. (λ) , and thus, by monotone convergence and the finiteness of $\tilde{P}_y$,

$$\sum_{k=-\infty}^{+\infty} p(\eta + 2k\pi)$$

is finite a.e. (λ) and, in fact, is the Radon-Nikodym derivative $d\tilde{P}_y/d\lambda$. It is then clear that

$$\frac{dP_y^n}{d\tilde{P}_y}(\eta) = \frac{dP_y^n/d\lambda\,(\eta)}{d\tilde{P}_y/d\lambda\,(\eta)} = \frac{p(\eta + 2n\pi)}{\sum_{k=-\infty}^{+\infty} p(\eta + 2k\pi)}$$

Finally, consider the set where $p_{y|\tilde{y}}(\xi|\tilde{\beta})$ is undefined -- i.e. where

$$\sum_{k=-\infty}^{+\infty} p(\tilde{\beta} + 2k\pi) = 0$$

But this set is a set of $\tilde{P}_y$-measure zero. Equation (94) follows immediately from (93) and the properties of the δ-function. ■

We make the comment that $\tilde{P}_y$ is the probability measure for the random variable $\tilde{y}$, and thus, a naive application of Bayes' rule yields

$$p_{y|\tilde{y}}(\xi|\tilde{\beta}) = \frac{p_{\tilde{y}|y}(\tilde{\beta}|\xi)\,p_y(\xi)}{p_{\tilde{y}}(\tilde{\beta})}$$

$$= \frac{\delta(\tilde{\beta} - (\xi \bmod 2\pi))\,p_y(\xi)}{\sum_{k=-\infty}^{+\infty} p_y(\tilde{\beta} + 2k\pi)}$$

We now consider a more general version of the problem stated at the beginning of this subsection. We assume that we have a probability space $(R^2, \mathscr{A}^2, P_{xy})$, where $\mathscr{A}^2$ is the Borel field of R^2. We define three random variables

$$x(\omega_1, \omega_2) = \omega_1$$

$$y(\omega_1, \omega_2) = \omega_2$$

$$\tilde{y}(\omega_1, \omega_2) = \omega_2 \bmod 2\pi$$

and the marginal distributions

$$P_x(A) = P_{xy}(A \times R^1) \qquad A \in \mathscr{A}^1$$

$$P_y(B) = P_{xy}(R^1 \times B) \qquad B \in \mathscr{A}^1$$

($\mathscr{A}^1 =$ Borel measurable subsets of R^1)

We let $\mathscr{A}_y$ = the minimum sub σ-algebra of $\mathscr{A}^2$, with respect to which y is measurable, and we define $\mathscr{A}_x$ and $\mathscr{A}_{\tilde{y}}$ analogously.

We wish to compute the conditional measure $P_{x|\tilde{y}}$. As before, we obtain a form for this measure that reflects our uncertainty as to the number of multiples of 2π that are "chopped off" of y in the process of observing $\tilde{y}$. To derive the desired result, we will need to consider two o her conditional measures, $P_{x|y,\tilde{y}}$ and $P_{x|y}$. Since $\mathscr{A}_{\tilde{y}} \subset \mathscr{A}_y$ (i.e. [illegible] a deterministic function of y), we have

$$[illegible](A|\beta, \tilde{\beta}) = P_{x|y}(A|\beta) . \qquad (95)$$

As before, we define the following measures on $([-\pi,\pi), \mathscr{S})$:

$$P_y^n(S) = P_y(S + 2n\pi)$$

$$\tilde{P}_y(S) = \sum_{n=-\infty}^{+\infty} P_y^n(S) \qquad S \varepsilon \mathscr{S}$$

Theorem 8: The conditional distribution $P_{x|\tilde{y}}(C|\tilde{\beta})$ is given by

$$P_{x|\tilde{y}}(C|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} \frac{dP_y^n}{d\tilde{P}_y}(\tilde{\beta})\, P_{x|y}(C|\tilde{\beta} + 2n\pi) \qquad (96)$$

Proof: It is easy to see from the previous results that the conditional probability

$$P_{y|\tilde{y}}(y = \mathscr{G}|\tilde{\beta}) = P_{x,y|\tilde{y}}(x \varepsilon R^1, \; y = \mathscr{G}|\tilde{\beta})$$

exists and is given by

$$P_{y|\tilde{y}}(y = \mathscr{G}|\tilde{\beta}) = \begin{cases} \dfrac{dP_y^k}{d\tilde{P}_y}(\tilde{\beta}) & \mathscr{G} = \tilde{\beta} + 2k\pi \\ 0 & \text{otherwise} \end{cases}$$

Using the properties of iterated expectations, [28], and equation (95)

$$P_{x|\tilde{y}}(C|\tilde{\beta}) = \mathscr{E}\{P_{x|y}(C|\mathscr{G})|\tilde{y} = \tilde{\beta}\}$$

$$= \int_{R^2} P_{x|y}(C|y(\omega_1,\omega_2) = \mathscr{G})\; dP_{x,y|\tilde{y}}(\omega_1,\omega_2|\tilde{y} = \tilde{\beta}) \; .$$

Since $P_{x|y}(C|\, y(\omega_1, \omega_2) = \mathscr{G})$ is independent of ω_1, we perform the integration with respect to ω_1 first

$$P_{x|\tilde{y}}(C|\tilde{\beta}) = \int_{R^1} P_{x|y}(C|\, y(\omega_2) = \mathscr{G})\, dP_{y|\tilde{y}}(\omega_2|\tilde{\beta})$$

$$= \sum_{n=-\infty}^{+\infty} \frac{dP_y^n}{d\tilde{P}_y}(\tilde{\beta})\; P_{x|y}(C|\, y = \tilde{\beta} + 2n\pi)$$

$$= \sum_{n=-\infty}^{+\infty} P(y = \tilde{\beta} + 2n\pi|\, \tilde{y} = \tilde{\beta})\; P_{x|y}(C|\, y = \tilde{\beta} + 2n\pi). \blacksquare$$

The following two corollaries solve the problem posed at the start of this subsection.

<u>Corollary 1:</u> Suppose x and v are independent real valued random variables, and define

$$y = m(x) + v$$
$$\tilde{y} = y \bmod 2\pi$$

where $m: R^1 \to R^1$ is measurable. Also, suppose $p_x(\alpha)$ and $p_v(\nu)$ are the probability densities for x and v respectively. Then a version of the probability density $p_{x|\tilde{y}}(\alpha|\tilde{\beta})$ is given by

$$p_{x|\tilde{y}}(\alpha|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} \frac{p_y(\tilde{\beta} + 2n\pi)}{p_{\tilde{y}}(\tilde{\beta})}\, p_{x|y}(\alpha|\tilde{\beta} + 2n\pi) \qquad (97)$$

$$= \sum_{n=-\infty}^{+\infty} \frac{p_{y|x}(\tilde{\beta} + 2n\pi|\alpha)\, p_x(\alpha)}{p_{\tilde{y}}(\tilde{\beta})} \qquad (98)$$

where

$$p_{y|x}(\tilde{\beta} + 2n\pi|\alpha) = p_v(\tilde{\beta} + 2n\pi - m(\alpha)) \tag{99}$$

$$p_y(\tilde{\beta} + 2n\pi) = \int_{-\infty}^{+\infty} p_{y|x}(\tilde{\beta} + 2n\pi|u)\, p_x(u)\, du \tag{100}$$

$$p_{\tilde{y}}(\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} p_y(\tilde{\beta} + 2n\pi) \tag{101}$$

<u>Proof:</u> Equation (97) follows from Theorem 8, the corollary to Theorem 7, and the observation that, if the measure $P_{x|y}(C|\beta)$ has a density with respect to Lebesque measure, then, from (96), so does $P_{x|\tilde{y}}(C|\tilde{\beta})$, and it is given by (97). Equations (99), (100), and (101) are immediate consequences of the definitions and the independence of x and v. Equation (98) follows from (97) and Bayes' rule. ■

<u>Corollary 2:</u> If x and v are normally distributed and m is linear ($m(x) = ax$), then $p_{x|\tilde{y}}(\alpha|\tilde{\beta})$ is expressible as the linear combination (102) of an infinite number of normal distributions, with weighting coefficients that are functions of the measurement and are given by

$$c_n(\tilde{\beta}) \triangleq \frac{p_y(\tilde{\beta} + 2n\pi)}{p_{\tilde{y}}(\tilde{\beta})} = \frac{N(\tilde{\beta} + 2n\pi;\ a\eta,\ a^2\gamma_1 + \gamma_2)}{\sum_{k=-\infty}^{+\infty} N(\tilde{\beta} + 2k\pi;\ a\eta,\ a^2\gamma_1 + \gamma_2)}$$

where

$$p_x(\alpha) = \frac{1}{\sqrt{2\pi\gamma_1}}\, e^{-\frac{(\alpha-\eta)^2}{2\gamma_1}} \triangleq N(\alpha;\ \eta,\ \gamma_1)$$

$$p_v(\nu) = \frac{1}{\sqrt{2\pi\gamma_2}} e^{-\nu^2/2\gamma_1} = N(\nu;\ 0, \gamma_2)$$

Proof: We will use the form of $p_{x|\tilde{y}}(\alpha|\tilde{\beta})$ given in (97). The additive properties of independent normally distributed random variables, [29], yields

$$p_y(\xi) = N(\xi;\ a\eta,\ a^2\gamma_1 + \gamma_2)$$

and therefore the equation for $c_n(\tilde{\beta})$ is correct. Then

$$p_{x|\tilde{y}}(\alpha|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} c_n(\tilde{\beta})\, p_{x|y}(\alpha|\tilde{\beta} + 2n\pi) \ . \tag{102}$$

But $p_{x|y}$ is the solution of a linear filtering problem, and therefore is a normal distribution. In fact

$$p_{x|y}(\alpha|\tilde{\beta} + 2n\pi) = N(\alpha;\ \eta_n, \gamma_3) \ ,$$

where

$$\gamma_3 = \frac{\gamma_1\gamma_2}{a^2\gamma_1 + \gamma_2}$$

$$\eta_n = \frac{\eta\gamma_2 + \gamma_1 a(\tilde{\beta} + 2n\pi)}{a^2\gamma_1 + \gamma_2} \qquad \blacksquare$$

That is, the n^{th} term in the series in (102) is evaluated by an optimal linear estimator which takes as its measurement $\tilde{\beta} + 2n\pi$. We also note that if the initial distribution $p_x(\alpha)$ is an infinite sum of

normal densities with means η_n, then the density $p_{x|\tilde{y}}(\alpha|\beta)$ is a doubly infinite sum of normal densities, with means computed by optimal linear estimators, the $(j,k)^{th}$ of which takes as its initial mean η_j and as its measurement $\tilde{\beta} + 2k\pi$. Again, the coefficients are nonlinear functions of the measurement.

Once having the solution $p_{x|\tilde{y}}(\alpha|\beta)$, we can compute $p_{\theta|\tilde{y}}(\tilde{\alpha}|\beta)$. If the hypotheses of Corollary 2 are satisfied, $p_{\theta|\tilde{y}}(\tilde{\alpha}|\beta)$ is an infinite sum of folded normal densities.

An interpretation of the form of the conditional density is readily available. The infinite summation is a result of the "mod 2π" ambiguity in the measurement. The n^{th} term in the sum is the linear result if the measurement were $y = \tilde{\beta} + 2n\pi$, while, as derived in Theorem 8 and its corollary, the coefficient $c_n(\tilde{\beta})$ is just $P(\tilde{y} = \beta + 2n\pi|\tilde{y} = \beta)$ -- i.e. it is related to the difference between y and $\tilde{y}$ expressed in multiples of 2π.

Thus, the terms corresponding to the more likely values of y -- the more likely number of multiples of 2π -- are more heavily weighted. Thus, one could consider approximating $p_{x|\tilde{y}}(\alpha|\beta)$ (and thus $p_{\theta|\tilde{y}}(\tilde{\alpha}|\beta)$) by a finite sum of normal distributions, where we must devise a procedure for deciding which terms to keep. Some work involving this type of approximation has been done by Buxbaum and Haddad, [30]. Such a procedure is certainly necessary if x is a random process instead of a random variable and we take a sequence of measurements, since, by a simple inductive argument, after M measurements our conditional density consists of M infinite sums of normal densities. Note that all the normal densities have the same variance.

In the particular case in which m is linear, we see that the conditional density is of the form

$$p(\theta) = \sum_{n=-1}^{\infty} C_n F(\theta; \eta_n, \gamma) \tag{103}$$

which is precisely the form studied in Section 2 (see equation (21)). Thus, the estimation and error analysis results of that section apply here. These results will also apply if we approximate (103) by a finite sum, and, since the truncation procedures of Buxbaum and Haddad and the estimation equations of Section 2 both lead to simple algorithms, this approach leads to easily implemented filter equations.

We remark that the appendix to this report contains results relating the discrete and continuous problems, by showing that as the time between measurements, Δt, becomes small, the terms in the conditional density corresponding to a nonzero number, n, of rotations between measurements go to zero exponentially in $1/\Delta t$ and in n^2. Thus we see that if the inter-measurement t ne is small, a rather crude truncation procedure -- one that keeps only a few terms, corresponding to one or two rotations -- will provide adequate accuracy.

In addition to the method of truncating the infinite series in some systematic manner, another suboptimal estimation scheme is suggested by the results of the a endix. Since for small Δt the difference between the continuous and discrete time solutions is small, why can't we use the continuous time results in designing a suboptimal discrete-time filter? That is, we can design the continuous time filter and use as

an input the discrete time measurements, which we hold constant over the interval between measurements.

We have not attempted to describe in detail the design of these various suboptimal estimation schemes, nor to analyze their performance, but rather we have only meant to indicate possible alternatives. Clearly further analysis and some simulation results are necessary before we can decide on the validity of these different approximations. The conceptual ideas behind these two basic methods are depicted in Figures 7 and 8 .

Analogous to the discussion at the end of Section 3, we can extend the results of the present section to problems on arbitrary abelian Lie groups. Let $\underline{x}$ be a random variable on R^{n+m} with probability density $p_{\underline{x}}(\alpha_1,\ldots,\alpha_{n+m})$, and consider the associated random variable, $\tilde{x}$, on $R^n \times (S^1)^m$ defined by the map from R^{n+m} into $R^n \times (S^1)^m$ given by

$$(x_1,\ldots,x_n,x_{n+1},\ldots,x_{n+m}) \rightarrow (x_1,\ldots,x_n,x_{n+1} \bmod 2\pi,\ldots, x_{n+m} \bmod 2\pi)$$

Then the density $p_{\tilde{x}}(\alpha_1,\ldots,\alpha_n,\tilde{\alpha}_1,\ldots,\tilde{\alpha}_m)$ is given by

$$p_{\tilde{x}}(\alpha_1,\ldots,\alpha_n,\tilde{\alpha}_1,\ldots,\tilde{\alpha}_m) =$$

$$\sum_{k_1=-\infty}^{+\infty} \cdots \sum_{k_m=-\infty}^{+\infty} p_{\underline{x}}(\alpha_1,\ldots,\alpha_n,\tilde{\alpha}_1 + 2k_1\pi,\ldots,\tilde{\alpha}_m + 2k_m\pi)$$

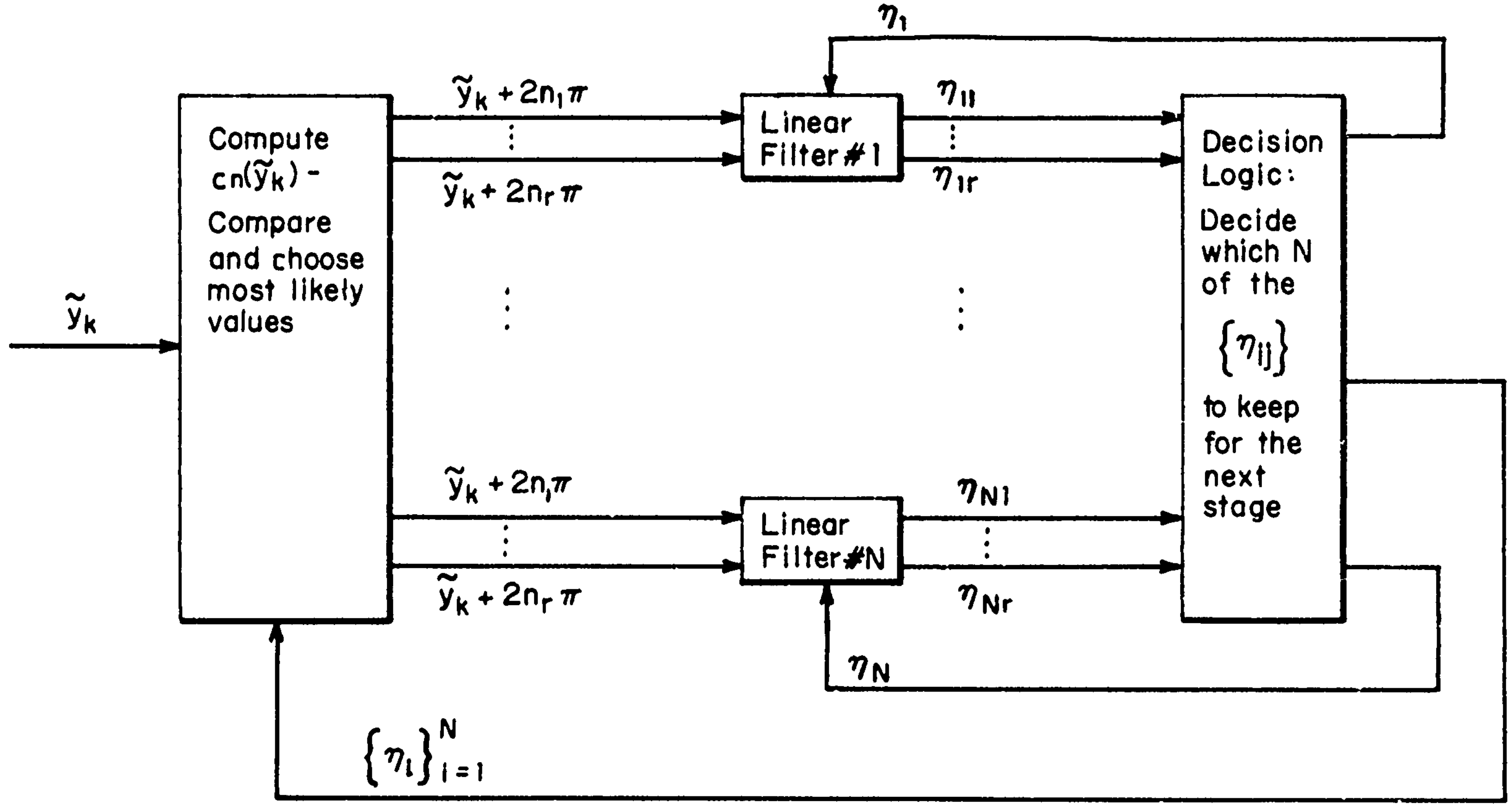

Figure 7 Conceptual Diagram for the Truncation Method for Suboptimal Discrete-Time Filtering (Each linear filter computes r estimates, using as data the given initial estimate η_i, the precomputed variance, and the r chosen input $\tilde{y}_k + 2n_j\pi$, $j = 1,\ldots,r$).

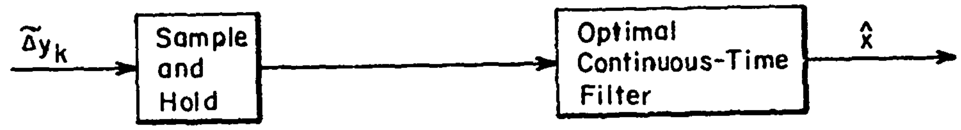

Figure 8: Illustrating the Concept of Using the Continuous-Time Filter to Approximate the Discrete-Time Filter

If p_x is a multidimensional normal density, then $p_{\tilde{x}}$ is called an (n, m) normal density -- n referring to the number of marginal densities which are normal and m to the number of folded normal marginal densities.

It is easy to see that minor changes in the arguments of this section lead to the following conclusion: let $C : R^{n+m} \rightarrow R^k$ be a linear map and w a k-dimensional normal random variable independent of $\underline{x}$, an (n+m)-dimensional normal random variable. Consider the random variable y defined by

$$y = C\underline{x} + w$$

and define the associated random variable $\tilde{y}$ by

$$\tilde{y}_i = y_i \qquad 1 \leq i \leq k_1$$

$$\tilde{y}_i = y_i \bmod 2; \qquad k_1 < i \leq k$$

Then the conditional density $f_{x|\tilde{y}}$ can be written as a $(k-k_1)$-times countably infinite sum of normal distributions, the $(r_1, \ldots, r_{k-k_1})^{th}$ of these being the linear result if

$$y_{k_1+i} = \tilde{y}_{k_1+i} + 2r_i\pi \qquad i = 1, \ldots, k-k_i \tag{104}$$

and the coefficient of this term is just the conditional probability for equation (104) to hold, given the random variable $\tilde{y}$.

4.2 The Discrete Measurement Problem Using Fourier Series

As was seen in subsection 2.2, Fourier series can be a useful tool. In this subsection we will use it to aid in analyzing a rather general discrete-time estimation problem on S^1. Again, we will consider the single measurement case. Extension to the multistage process with measurement noise independent from stage to stage is immediate.

We consider the problem of taking a measurement of a random variable, θ, on the circle with a priori density

$$p_\theta(\xi) = \frac{1}{2\pi} + \sum_{n=1}^{\infty} a_n \sin n\xi + b_n \cos n\xi$$

We assume that we take a single (possibly nonlinear) measurement, y, of θ, and that the conditional density $p_{y|\theta}(\beta|\xi)$ exists. Considering this as a function of ξ for fixed β, we must have

$$p_{y|\theta}(\beta|\xi + 2\pi) = p_{y|\theta}(\beta|\xi) .$$

Thus, we can write $p_{y|\theta}(\beta|\xi)$ in Fourier series form in ξ for fixed β

$$p_{y|\theta}(\beta|\xi) = d_0(\beta) + \sum_{n=1}^{\infty} c_n(\beta) \sin n\xi + d_n(\beta) \cos n\xi$$

where the c_n's and d_n's are functions of β. An application of Bayes' rule yields the Fourier series form for the conditional density $p_{\theta|y}(\xi|\beta)$

$$p_{\theta|y}(\xi|\beta) = \frac{1}{2\pi} + \sum_{n=1}^{\infty} a_n(\beta) \sin n\xi + b_n(\beta) \cos n\xi \tag{105}$$

where

$$a_n(\beta) = \frac{\alpha_n(\beta)}{2\pi\, c(\beta)} \;;\; b_n(\beta) = \frac{\mathscr{G}_n(\beta)}{2\pi\, c(\beta)} \tag{106}$$

with

$$c(\beta) = \frac{1}{2\pi} p_y(\beta) = \frac{d_0(\beta)}{2\pi} + \frac{1}{2} \sum_{n=1}^{\infty} \left[a_n c_n(\beta) + b_n d_n(\beta)\right] \tag{107}$$

$$\alpha_k(\beta) = a_k d_0(\beta) + \frac{c_k(\beta)}{2\pi} + \frac{1}{2} \sum_{n=1}^{k-1} \left[a_n d_{k-n}(\beta) + b_n c_{k-n}(\beta)\right]$$

$$+ \frac{1}{2} \sum_{n=1}^{\infty} \left\{\left[a_{n+k} d_n(\beta) + b_n c_{n+k}(\beta)\right] - \left[a_n d_{n+k}(\beta) + b_{n+k} c_n(\beta)\right]\right\}$$

(108)

$$\mathscr{G}_k(\beta) = b_k d_0(\beta) + \frac{d_k(\beta)}{2\pi} + \frac{1}{2}\sum_{n=1}^{k-1}\left[b_n d_{k-n}(\beta) - a_n c_{k-n}(\beta)\right]$$

$$+\frac{1}{2}\sum_{n=1}^{\infty}\left\{\left[a_{n+k}\, c_n(\beta) + b_n d_{n+k}(\beta)\right] + \left[a_n c_{n+k}(\beta) + b_{n+k} d_n(\beta)\right]\right\} \quad (109)$$

Note that the equations for c, α_k, and $\mathscr{G}_k$ are bilinear in the Fourier coefficients of $p_\theta(\xi)$ and $p_{y|\theta}(\beta|\xi)$, and one should note the marked similarity in the structure of (108) and (109). Thus, the computation of $p_{\theta|y}(\xi|\beta)$ involves the (in general nonlinear) computation of the coefficients $\{c_n(\beta)\}$ and $\{d_n(\beta)\}$ and the evaluation of the bilinear equations (107), (108), and (109).

The form of these conditional density equations suggests a truncation of the Fourier series for p_θ and $p_{y|\theta}$, which leads to finite sums in (107) through (109), however if we retain the first N modes of p_θ and the first M modes of $p_{y|\theta}$, then $p_{\theta|y}$ will have terms up to the $(N+M)^{th}$ mode. Thus, to keep the necessary memory in a multistage process from growing in this manner, it becomes necessary to devise techniques for sequentially truncating the conditional density for θ. We will not treat this problem in detail, but will make some general comments. In general, just keeping the first N modes of $p_{\theta|y}$ is not an acceptable method, since we require that the truncated density be nonnegative everywhere. However, if, for instance $p_{\theta|y}$ is continuous, the coefficients fall off as $\frac{1}{N^2}$, and thus, for any given $\epsilon > 0$, we can choose N sufficiently large so that, if we keep the first N modes, the truncated density will be bounded below

by $-\epsilon$. Alternatively, if $\tilde{p}$ represents the truncated version of $p_{\theta|y}$ obtained by keeping only the first N modes, and we then define

$$\tilde{\tilde{p}} = \max(0, \tilde{p})$$

we can take the Fourier coefficients of $\tilde{\tilde{p}}$ as the coefficients of our approximation to $p_{\theta|y}$.

If we were to use the straightforward method of truncating the Fourier series for $p_{\theta|y}$, equations (107) through (109) can be written

$$g(\beta) = A(\beta)h \tag{110}$$

where h is the vector whose elements are the Fourier coefficients of $p_\theta(\xi)$, $g(\beta)$ contains $c(\beta)$ and the $\alpha_k(\beta)$'s and $\mathscr{G}_k(\beta)$'s and $A(\beta)$ is a $(2N+1) \times 2N$ matrix (assuming we keep N modes of p_θ and $p_{\theta|y}$) whose elements are the Fourier coefficients of $p_{y|\theta}$. The structure of (107), (108), and (109) is reflected in A and may lead to efficient methods for evaluating (110) .

Finally, we note that this approach is extremely general, in that the only restriction on the form of the measurement is that the conditional density $p_{y|\theta}$ exist. For example, in addition to measurements such as

$$y = (\theta + v) \bmod 2\pi$$

which are considered in subsection 4. 1 , using the Fourier series approach we can also consider measurements such as

$$y = \sin\theta + v \ .$$ [1]

[1] It has recently been pointed out to the authors that Fourier analysis results for this particular measurement form were presented in ref. 47.

5. Applications Including FM-AM Demodulation

The previous sections have described and analyzed a class of mathematical models for which relatively simple optimal filter have been obtained. The problems considered include many inherently nonlinear ones. These results are interesting in that they provide a new way of introducing randomness into the system equations in such a manner as to lead to simple synthesis procedures for optimal estimation.

Among the potential areas of application are FM demodulation, AM demodulation, combined AM-FM demodulation, optical communication, frequency stability, and gyroscopic analysis.

The usual mathematical models for the received FM signal are (refs. 31, 32, 33)

$$r(t) = A \cos (\omega_0 t + \int_0^t x(s)ds) + N_1(t) \tag{111}$$

or

$$r(t) = A \cos (\omega_0 t + \int_0^t x(s)ds + N_2(t)) \tag{112}$$

where N_1 and N_2 are noise processes (here assumed to be Brownian motion processes), and $\dot{x}$ is the signal.

It is the mathematical model (112) that we will consider in this section. More detailed descriptions of and other analyses using this model can be found in references 31-39 and 45.

We remark that there are techniques for determining $\sin (\omega_0 t + \int_0^t x(s)ds + N_2(t))$ from $r(t)$. Using the notation of the preceding sections, we take as our observation the 2 x 2 orthogonal matrix

$$Z(t) = \begin{bmatrix} \cos(\omega_0 t + \int_0^t x(s)ds + N_2(t)) & \sin(\omega_0 t + \int_0^t x(s)ds + N_2(t)) \\ -\sin(\omega_0 t + \int_0^t x(s)ds + N_2(t)) & \cos(\omega_0 t + \int_0^t x(s)ds + N_2(t)) \end{bmatrix}$$

Then we can apply the estimation results of previous sections to obtain an optimal estimate for the signal $x(t)$ (see Lemma 3).

If the signal $x(t)$ is a linear diffusion process, the optimal demodulation equations take a particularly simple form. Also, if we have a multi-channel FM system, we can model it à la subsection 3.5 and use the results on filtering in abelian Lie groups to design an optimal frequency demodulator.

The theory developed in this report also has possible applications in AM modulation, joint AM-FM modulation (ref. 33, p. 628), and optical communication. The Lie group of interest in these cases is $C - \{0\}$ -- the set of nonzero complex numbers with complex multiplication as the group operation. Its (real) Lie algebra can be identified as R^2, and the map $\exp: R^2 \longrightarrow C - \{0\}$ is defined by

$$\exp(x_1, x_2) = e^{x_1 + ix_2} \quad . \tag{113}$$

We note that $C - \{0\} \approx R^1 \times S^1$ via the identification

$$(r, \theta) \longrightarrow e^{r + i\theta} \qquad r \in R, \ \theta \in [-\pi, \pi)$$

Thus S^1 is the subgroup of $C - \{0\}$ consisting of all complex numbers of length one, and its Lie algebra is the subalgebra of R^2 obtained by requiring $x_1 = 0$. We note that this representation of S^1 could have been used in the preceding sections, instead of the 2×2 orthogonal matrices.

Also, we see from (113) that x_1 controls the amplitude of $\exp(x_1, x_2)$, while x_2 controls the phase.

Now suppose that we have a continuous signal process on R^2

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

We define our measurement process, z(t), as follows:

$$dy(t) = x(t)dt + dv(t) \tag{114}$$

$$z(t) = \exp(y_1(t), y_2(t)) = e^{y_1(t) + i y_2(t)} \tag{115}$$

where $v' = (v_1, v_2)$ is a 2-dimensional Brownian motion process, independent of x.

This problem clearly fits into the framework discussed in Section 3, and thus can be solved by the methods described previously -- i. e. knowledge of $z(s)$, $s \leq t$ is equivalent to knowledge of $y(s)$, $s \leq t$. In fact, we can express $dy(t)$ in terms of $z(t)$ and $dz(t)$ with the aid of the Ito differential rule:

$$dz(t) = (dy_1(t) + idy_2(t))z(t) + \left| \frac{q_{11}(t) + 2iq_{12}(t) - q_{22}(t)}{2} \right| z(t)dt$$

where

$$E(dv(t)dv'(t)) = Q(t)dt$$

$$Q(t) \begin{vmatrix} q_{11}(t) & q_{12}(t) \\ q_{12}(t) & q_{22}(t) \end{vmatrix}$$

Thus

$$dy_1(t) = \mathrm{Re}\left\{\frac{dz(t)}{z(t)}\right\} - \frac{q_{11}(t) - q_{22}(t)}{2}\,dt$$

$$dy_2(t) = \mathrm{Im}\left\{\frac{dz(t)}{z(t)}\right\} - q_{12}(t)dt$$

Also note that, if we assume $y(0) = 0$, we have

$$y(t) = \int_0^t x(s)ds + v(t)$$

and thus

$$z(t) = \left[e^{v_1(t) + iv_2(t)}\right]\left[e^{\int_0^t [x_1(s) + ix_2(s)]ds}\right] \tag{116}$$

We then see that our signal is both amplitude and frequency modulated, and the noise enters multiplicatively and is a complex lognormal process (ref. 40).

Thus, equation (116) yields a message model for a joint AM-FM modulation system, for which there is a simple optimal estimator. The AM case is obtained by setting $x_2 \equiv v_2 \equiv 0$. We note that our AM modulation is not the usual one -- actually $x_1(t)$ is more like an amplitude <u>rate</u> modulating signal. However, if we let $x_1(t) = (d/dt)\,\tilde{x}_1(t)$, where $\tilde{x}_1(t)$ is the actual signal we want transmitted, we have that the amplitude modulation is (assuming $x_1(0) = 0$ and $\tilde{x}_1$ is deterministic).

$$e^{\int_0^t x_1(s)\,ds} \qquad e^{\tilde{x}_1(t)}$$

Thus

$$dy_1(t) = \mathrm{Re}\left\{\frac{dz(t)}{z(t)}\right\} - \frac{q_{11}(t) - q_{22}(t)}{2}\, dt$$

$$dy_2(t) = \mathrm{Im}\left\{\frac{dz(t)}{z(t)}\right\} - q_{12}(t)dt$$

Also note that, if we assume $y(0) = 0$, we have

$$y(t) = \int_0^t x(s)ds + v(t)$$

and thus

$$z(t) = \left[e^{v_1(t) + iv_2(t)}\right]\left[e^{\int_0^t [x_1(s) + ix_2(s)]ds}\right] \tag{116}$$

We then see that our signal is both amplitude and frequency modulated, and the noise enters multiplicatively and is a complex lognormal process (ref. 40).

Thus, equation (116) yields a message model for a joint AM-FM modulation system, for which there is a simple optimal estimator. The AM case is obtained by setting $x_2 \equiv v_2 \equiv 0$. We note that our AM modulation is not the usual one -- actually $x_1(t)$ is more like an amplitude <u>rate</u> modulating signal. However, if we let $x_1(t) = (d/dt)\,\tilde{x}_1(t)$, where $\tilde{x}_1(t)$ is the actual signal we want transmitted, we have that the amplitude modulation is (assuming $x_1(0) = 0$ and $\tilde{x}_1$ is deterministic).

$$e^{\int_0^t x_1(s)\, ds} \qquad e^{\tilde{x}_1(t)}$$

or, if $\tilde{x}_1$ is deterministic and differentiable, and we let $x_1(t) \frac{(d/dt)\tilde{x}_1(t)}{\tilde{x}_1(t)}$, then

$$e^{\int_0^t x_1(s)ds} = \tilde{x}_1(t) / \tilde{x}_1(0)$$

(assuming $\tilde{x}_1 \geq 0$). Also note that in the AM case $(x_2 \equiv 0)$ we can include $v_2(t)$ as a random phase, and in the FM case $(x_1 \equiv 0)$ we can include $v_1(t)$ as a random amplitude.

In optical communication theory, variations in the transmission medium -- e. g. turbulence in the atmosphere -- cause variations in the refractive index of the air. This disturbance can be modeled (ref. 40) as a lognormal noise process which multiplies the signal. In this case, this analysis (equations (113) through (116)) may prove to be helpful in the design of good receivers. In particular, these results may be useful in the case of spatially uniform noise, and, in addition, we can treat the problem with real and imaginary parts of the noise process dependent on each other ($q_{12}(t) \not\equiv 0$, see ref. 40).

The problem of frequency stability (refs. 41, 42, 43) is another area of application of the results of this report. This problem involves devices, such as oscillators and extremely accurate clocks, in which we wish to measure deviations of the operating frequency from some ideal or nominal frequency. In other words, we have a signal of the form

$$e^{i(\omega_0 t + \int_0^t x(s)ds)}$$

where ω_0 is the fixed, ideal frequency and $x(s)$ is the random, time-varying deviation of the actual frequency from the ideal frequency. The

problem is to devise a measurement and estimation system to determine the deviation $x(t)$.

There are various types of measurement processes discussed in the literature (refs. 41, 42, 43). One of the most widely mentioned involves the multiplication of the signal by the output of a second oscillator and the measurement of the beat frequency. That is, if the signal from the second oscillator is

$$e^{i(\omega_1 t - v(t))}$$

where ω_1 is a fixed frequency, close to ω_0, and $v(t)$ is a random deviation from ω_1, our measurement essentially is

$$e^{i[(\omega_0 - \omega_1)\, t + \int_0^t x(s)ds + v(t)]}$$

If we assume that v is a Brownian motion process independent of x, and if we subtract off the known term $(\omega_0 - \omega_1)t$, we are left with the observation equations

$$dz(t) = x(t)dt + dv(t)$$

$$Z(t) = e^{iz(t)}$$

which is precisely the form considered in this report. Further if we model x as a Brownian motion process or a linear diffusion process, we can use the optimal filtering equations of subsection 3.3.

A final area of application is in the estimation of the angular position of a body spinning about a given axis. If we consider the single-degree-of-freedom integrating gyroscope, (ref. 44, pp. 104-105), we note that the

output of this device is an angle -- essentially the shift in orientation of the gyro from some reference position. The orientation of the gyro is determined by the integral of the angular velocity acting on the gyro about some fixed axis. Noise in the system is modeled as gyro drift -- an error in the angular velocity detected by the device. Using this model for the dynamics, the estimation results of this report can be used to design a system to estimate the actual angular velocity.

6. Conclusions

In this report, a class of estimation problems on the unit circle is formulated and resolved. Both the continuous time and discrete time estimation problems are considered. The signal and observation processes on the circle are constructed by taking the projection modulo 2π of the corresponding standard 1-dimensional processes. The stochastic differential equations which govern their evolution are bilinear in form. The observational noise can be viewed as entering multiplicatively.

Error criteria, probability distributions, and optimal estimates on the circle are studied. In particular, various properties of the folded normal density in connection with estimation are discussed in detail.

An effective synthesis procedure for continuous time estimation is provided. The measurement data is first processed through a nonlinear transformation. The transformed process then goes through an ordinary estimator, such as the Kalman-Bucy filter. After another nonlinear processing of the output of the ordinary estimator, the desired estimate is yielded. Filtering, smoothing and prediction can all be treated in this manner, and its generalization to estimation on an arbitrary abelian Lie group finds no difficulty.

In addition, the discrete time problem was studied, and an intrinsic difference between the continuous and discrete problems was discussed. This difference stems from the loss of information between the discrete measurements. Unlike the vector space case, this loss of information causes the expression for the conditional probability distribution to be rather cumbersome. Although suboptimal estimators can be obtained from the results of Section 4 by careful examination of the form of the equations,

the increasing complexity of these equations with each additional measurement has prevented the authors from deriving recursive equations for the optimal estimate.

Applications to AM and FM demodulation, optical communication, frequency stability, and fixed axis rotation problems have been described. These practical problems provide physical justification for the proposed mathematical formulation. The application of the mathematical results of this paper is seen to lead to neat solution and easy implementation in these practical situations.

Acknowledgements

The authors wish to acknowledge Professor Roger W. Brockett of Harvard University, who provided the motivation and direction for this work, and Professor Jan C. Willems and Professor John J. Deyst of M.I.T. for their criticism and advice during the preparation of this report.

REFERENCES

1. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, Ser. D, J. Basic Eng., Vol. 82, 1960, pp. 35-45.

2. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," Trans. ASME, Ser. D, J. Basic Eng., 1961, pp. 95-108.

3. J. T. Lo, "Finite Dimensional Sensor Orbits and Optimal Nonlinear Filtering," Tech. Report, USCAE 114, August 1969 (to appear IEEE Trans. on Information Theory, Sept. 1972).

4. P. Levy, "L'Addition des Variables Aléatoires Définies sur une Circonference," Bull. Société Math. France, Vol. 67, 1939, pp. 1-41.

5. F. Perrin, "Étude Mathematique du Mouvement Brownien de Rotation," Ann. Ecole Norm. Sup., Vol. 45, 1928, pp. 1-51.

6. A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

7. K. Itô, "On a Formula Concerning Stochastic Differentials," Nagoya Math. J., Vol. 3, 1951, pp. 55-65.

8. G. H. Hardy, J. E. Littlewood, G. Pólya, Inequalities, Cambridge, 1952.

9. G. Szegö and G. Pólya, Isoperimetric Inequalities in Mathematical Physics, Princeton, 1951.

10. S. Sherman, "Non Mean-Square Error Criteria," IRE Trans. on Inf. Theory, Vol. 4, 1958, pp. 125-126.

11. S. Sherman, "A Theorem on Convex Sets with Applications," Ann. Math. Stat., Vol. 26, 1955, pp. 763-767.

12. R. Bellman, A Brief Introduction to Theta Functions, Holt, Rinehart, and Winston, New York, 1961.

13. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, Amer. Ed., The MacMillan Co., New York, 1945.

12. J. L. Brown, "Asymmetric Non-Mean-Square Error Criteria," IRE Trans. on Aut. Cont., Vol. 7, 1962, pp. 64-66.

15. T. T. Kadcta and L. A. Shepp, "Conditions for Absolute Continuity between a Certain Pair of Probability Measures," Z. Wahrscheinlichkeitstheorie verw. Geb. 16, pp. 250-260, 1970.

16. J. T. Lo, "A General Bayes Rule and A General Bucy Representation Theorem," submitted for publication.

17. J. T. Lo, "On Optimal Nonlinear Estimation -- Part I: Continuous Observation," Proc. of the 8th Annual Allerton Conference on Circuits and Systems, 1970; to appear in the Journal of Information Sciences.

18. P. A. Frost and T. Kailath, "An Innovations Approach to Least-Squares Estimation -- Part III: Nonlinear Estimation in White Gaussian Noise," The IEEE Transactions on Automatic Control, June, 1971.

19. R. L. Stratonovich, "Conditional Markov Processes," Theory of Probability and its Applications, Vol. 5, 1962.

20. H. J. Kushner, "On The Dynamic Equations of Conditional Probability Density Functions, with Applications to Optimal Stochastic Control Theory," J. of Math. Anal. Appl., Vol. 8, 1964.

21. K. Ito and H. P. McKean, Jr., Diffusion Processes and Their Sample Paths, Academic Press, New York, 1965.

22. T. Kailath and P. Frost, "An Innovations Approach to Least Squares Estimation -- Part II: Linear Smoothing in Additive White Noise," IEEE Trans. Automatic Control, Vol. AC-13, Dec. 1968, pp. 655-660.

23. F. W. Warner, Foundations of Differentiable Manifolds and Lie Groups, Scott, Foresman and Co., Glenview, Ill. 1971.

24. W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1966.

25. P. R. Halmos, Measure Theory, Van Nostrand Reinhold Co., New York, 1950.

26. J. L. Doob, Stochastic Processes, John Wiley and Sons, New York, 1953.

27. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. II, John Wiley and Sons, New York, 1966.

28. J. Neveu, Mathematical Foundations of the Calculus of Probability, Holden-Day, 1965.

29. K. S. Miller, Multidimensional Gaussian Distributions, Wiley, New York, 1964.

30. P. J. Buxbaum and R. A. Haddad, "Recursive Optimal Estimation for a Class of Nongaussian Processes," Proc. Symp. on Computer Processing in Comm., Polytechnic Institute of Brooklyn, April 8-10, 1969.

31. D. L. Snyder, The State-Variable Approach to Continuous Estimation with Applications to Analog Communication Theory, M. I. T. Press, Camb., Mass., 1969.

32. J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering, John Wiley and Sons, New York, 1965.

33. D. Middleton, An Introduction to Statistical Communication Theory, McGraw-Hill, New York, 1960.

34. W. A. Edson, "Noise in Oscillators," Proc. of the IRE, Vol. 48, No. 8, 1960, pp. 1454-1466.

35. J. A. Develet, "Threshold Criterion for Phase-Lock Demodulation," Proc. of the IEEE, Vol. 51, No. 2, Feb. 1963, pp. 349-356.

36. R. A. LaPlante, "Development of a Low Noise X-Band CW Klystron Power Oscillator," Trans of the IRE, PGED, Vol. ED-1, No. 4, Dec. 1954, pp. 99-106.

37. D. Middleton, "Theory of Phenomenological Models and Measurement of the Fluctuating Output of CW Magnetrons," Trans of the IRE, PGED, Vol. ED-1, No. 4, Dec. 1954, pp. 56-90.

38. J. L. Stewart, "Theory of Frequency Modulation Noise in Tubes Employing Phase Focusing," J. of Applied Physics, Vol. 26, No. 4, April 1955, pp. 409-413.

39. A. J. Viterbi, Principles of Coherent Communication, McGraw-Hill, New York, 1966.

40. E. V. Hoversten, R. O. Harger, and S. J. Halme, "Communication Theory for the Turbulent Atmosphere," Proc. of the IEEE, Vol. 58, No. 10, Oct. 1970, pp. 1626-1650.

41. D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. of the IEEE, Vol. 54, No. 2, Feb. 1966, pp. 221-230.

42. J. A. Barnes, "Atomic Timekeeping and the Statistics of Precision Signal Generators," Proc. of the IEEE, Vol. 54, No. 2, Feb. 1966, pp. 207-220.

43. L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," Proc. of the IEEE, Vol. 54, No. 2, Feb. 1966, pp. 136-154.

44. W. Wrigley, W. Hollister, and W. Denhard, Gyroscopic Theory, Design, and Instrumentation, The M.I.T. Press, Cambridge, Mass., 1969.

45. M. G. Crosby, "Frequency-Modulation Noise Characteristics," Proc. of the IRE, Vol. 25, April, 1937, pp. 472-514.

46. Y.H. Ku, L. P. Sheporaitis, "Global Properties of Diffusion Processes on Cylindrical Type Phase Space," J. Of Franklin Institute, Vol. 289, No. 2, Feb. 1970, pp. 1087-1103.

47. A. J. Mallinckrodt, R. S. Bucy, S. Y. Cheng, "A Design Study for an Optimal Non-Linear Receiver/Demodulator," Final Report N. A. S. A. Contract No. NAS5-10789, 1970.

APPENDIX

Some Limiting Arguments Relating the Discrete and Continuous Problems

We have seen in Section 4 that the ambiguity concerning the number of rotations leads to equations for the conditional density that involve infinite sums. Intuitively, if we observe the process continuously, this ambiguity should disappear -- assuming the random processes involved are continuous. From the rigorous arguments of Section 3, we have seen that this is the case -- i.e. in the limit we know $dy(t)$, not just $dy(t) \bmod 2\pi$. We can also see this by examining the discrete approximation to the continuous problem.

Our discrete equations are

$$\begin{aligned}\widetilde{\Delta} y_k &= (\Delta y_k) \bmod 2\pi \\ &= [m(x_k, k\Delta t)\,\Delta t + \sqrt{q(k\Delta t)}\;\Delta w_k] \bmod 2\pi\end{aligned}$$

where $x_k = x(k\Delta t)$, and $x(t)$ is a continuous process, independent of the Brownian motion process $w(t)$. We also assume $q(t)$ is continuous, $m(x,t)$ is measurable in x for all t and continuous in t for all x.

We wish to examine the effect of one additional such measurement at time t, in terms of the size of Δt. Thus, we assume we have computed

$$p_{x(t)}(\alpha) \triangleq p_{x(t)}(\alpha \mid \text{past measurements}) \tag{117}$$

and that we take the measurement

$$\tilde{\Delta} y_t = (\Delta y_t) \bmod 2\pi$$

$$= (m(x(t), t)\,\Delta t + \sqrt{q(t)}\,\Delta w_t) \bmod 2\pi$$

As indicated in equation (117) we will suppress all conditioning on past measurements. Thus, we wish to compute $p_{x(t)|\tilde{\Delta}y_t}(\alpha|\tilde{\beta})$ in terms of $p_{x(t)}(\alpha)$ and the new information $\tilde{\Delta}y_t$. (Here $\tilde{\beta}$ is the observed value of $\tilde{\Delta}y_t$).

Using the discrete measurement formulae, we have

$$p_{x(t)|\tilde{\Delta}y_t}(\alpha|\tilde{\beta}) = \sum_{n=-\infty}^{+\infty} c_n(\tilde{\beta})\, p_{x(t)|\Delta y_t}(\alpha|\tilde{\beta} + 2n\pi) \tag{118}$$

where we have the explicit formula

$$c_n(\tilde{\beta}) = \frac{\int_{-\infty}^{+\infty} N(\tilde{\beta} + 2n\pi - m(u,t)\,\Delta t;\ 0, q(t)\,\Delta t)\, p_{x(t)}(v)\, dv}{\sum_{r=-\infty}^{+\infty} \int_{-\infty}^{+\infty} N(\tilde{\beta} + 2r\pi - m(u,t)\,\Delta t;\ 0, q(t)\,\Delta t)\, p_{x(t)}(u)\, du}$$

$$= \frac{\int_{-\infty}^{+\infty} \exp - \frac{1}{2q(t)\,\Delta t}\{\tilde{\beta} + 2n\pi - m(u,t)\,\Delta t\}^2\, p_{x(t)}(u)\, du}{\sum_{r=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp - \frac{1}{2q(t)\,\Delta t}\{\tilde{\beta} + 2r\pi - m(u,t)\,\Delta t\}^2\, p_{x(t)}(u)\, dv} \tag{119}$$

Examining this expression, we see that the numerator contains a term

$$\exp - \frac{2n^2\pi^2}{q(t)\,\Delta t}$$

which is $o(\Delta t^{\ell})$ $\forall \ell \geq 0$ if $n \neq 0$. Thus, one sees that for small Δt, the probability that Δy_t and $\widetilde{\Delta} y_t$ differ by a nonzero multiple of 2π appears to go to zero quite fast as $\Delta t \to 0$. To make a precise statement concerning this, we must make some technical assumptions:

(1) The probability density for $x(t)$ conditioned on the past measurements, $p_{x(t)}(\alpha)$, exists.

(2) The conditional density for $x(t)$, if we were to measure Δy_t (<u>not</u> $\widetilde{\Delta} y_t$), $p_{x(t)|\Delta y_t}(\alpha|\beta)$ exists and is bounded uniformly for all α and β.

(3) We have the following bound:

$$\int_{-\infty}^{+\infty} e^{-\alpha^2 m^2(u,t) + \xi m(u,t)} \; p_{x(t)}(u)\, du \leq K(\alpha^2)\, e^{k(\alpha^2)\xi^2} \tag{120}$$

where $K(\alpha^2)$, $k(\alpha^2)$ are bounded for $\alpha^2 \in [0,\gamma]$, for some $\gamma > 0$.

We can now prove the following

<u>Theorem 9:</u> If the assumptions above hold, we have the following relationship among the c_n's:

$$\sum_{n\neq 0} \frac{c_n(\tilde{\beta})}{c_0(\tilde{\beta})} = o(\Delta t^{\ell}) \qquad \forall \ell \geq 0 \tag{121}$$

$$\left\| p_{x(t)|\Delta y_t}(\alpha|\tilde{\beta}) - p_{x(t)|\widetilde{\Delta} y_t}(\alpha|\tilde{\beta}) \right\|_{L^\infty} = o(\Delta t^{\ell}) \quad \forall \ell \geq 0 . \tag{122}$$

We will need the following technical lemma:

Lemma 5: Let $y(t)$ be a continuous, real-valued function of time, and define

$$\tilde{\Delta} y_t(s) = (y(s) - y(t)) \bmod 2\pi \qquad (s > t)$$

Let q be a positive constant, $\rho(x)$ a measurable real valued function on R^1, $h \geq 0$, an element of $L^1(R)$ such that $\|h\|_{L^1} > 0$, and $\{s_n\}_{n=1}^{\infty}$ a sequence of real numbers decreasing to t. Then there exists an integer N_0, such that

$$\int_{-\infty}^{+\infty} \exp - \frac{1}{2q} \{ \rho(x)(s_n - t) - 2\tilde{\Delta} y_t(s_n)\, \rho(x) \}\, h(x)\, dx$$

$$\geq \frac{1}{2} \|h\|_{L^1} \qquad \forall\, n \geq N_0 \tag{123}$$

Proof: Let $F_n(x)$ be the integrand of the left-hand side of (123). Then, for fixed x

$$\lim_n F_n(x) = h(x)$$

By Fatou's Lemma ([24], [25])

$$\liminf_n \int F_n\, dx \geq \int \lim_n F_n\, dx$$

$$= \|h\|_{L^1}$$

Choose N_0, such that for $m \geq N_0$

$$\int F_m \, dx \geq \liminf_n \int F_n \, dx - \frac{1}{2} \|h\|_{L^1}$$

Then

$$\int F_m \, dx \geq \frac{1}{2} \|h\|_{L^1} \qquad \forall \, m \geq N_0$$

■

Proof of Theorem 9: The proof of this result requires some straightforward but rather lengthy computations. Thus, we shall only sketch the proof, leaving the details to the interested reader.

Consider the infinite sum

$$\tilde{d}(\beta) \triangleq \sum_{n \neq 0} \int_{-\infty}^{+\infty} \exp - \frac{1}{2q(t)\,\Delta t} \{2\tilde{\beta}(2n\pi - m(u,t)\,\Delta t)$$

$$+ (2n\pi - m(u,t))^2\} \, p_{x(t)}(u) \, du$$

Using (120), we have

$$\tilde{d}(\beta) \leq \sum_{n \neq 0} \exp - \frac{n\pi(2n\pi - 2\tilde{\beta})}{q(t)\,\Delta t} \, K\left(\frac{\Delta t}{2q(t)}\right) \exp\left\{k\left(\frac{\Delta t}{2q(t)}\right)\left(\frac{2n\pi + \beta}{q(t)}\right)^2\right\}$$

We now note that $w(t)$ is a continuous random process, and, therefore, we assume that we are given a continuous sample path, $w^0(t)$. We then choose a $\delta > 0$, $K_0 > 0$, $k_0 > 0$, such that

$$K\left(\frac{\Delta t}{2q}\right) \leq K_0$$

$$k_0\left(\frac{\Delta t}{2q}\right) \le k_0$$

$$|\tilde{\Delta} y_t^0| \le \pi/2$$

for all $\Delta t < \delta$ (here $\tilde{\Delta} y_t^0$ is the value for the particular sample path $w^0(t)$ selected). Then, for $\Delta t < \delta$.

$$d(\tilde{\Delta} y_t^0) \le K_0 \sum_{n \neq 0} \exp -\left\{\frac{(2n^2 - |n|)\pi^2}{q(t)\,\Delta t}\right\} \exp k_0 \left\{\frac{(2|n| + 1)}{q(t)} \pi\right\}^2 \tag{124}$$

and for $\Delta t < \min(\delta, q(t)/2k_0)$, the right hand side of (124) is finite.

Examining equation (119), we can write

$$\frac{1}{\Delta t^{\ell}} \sum_{n \neq 0} \frac{c_n(\tilde{\Delta} y_t^0)}{c_0(\tilde{\Delta} y_t^0)} =$$

$$= \frac{1}{\Delta t^{\ell}} \frac{d(\tilde{\Delta} y_t^0)}{\displaystyle\int_{-\infty}^{+\infty} \exp - \frac{1}{2q(t)} \{m(u,t)\,\Delta t - 2\tilde{\Delta} y_t^0\, m(u,t)\}\; p_{x(t)}(u)\, du}$$

Taking a sequence $\{\Delta t_r\}_{r=1}^{\infty}$ decreasing to zero and using Lemma 5, we see that there is an R_0 such that

$$\int_{-\infty}^{+\infty} \exp - \frac{1}{2q(t)} \{m(u,t)\,\Delta t_r - 2\tilde{\Delta} y_t^0(r)\, m(u,t)\}\; p_{x(t)}(u)\, du \ge \frac{1}{2}$$

$$\forall\, r \ge R_0$$

where

$$\widetilde{\Delta}y_t^0(r) = (m(x(t),t)\,\Delta t_r + q(t)\,\Delta_r w_t^0) \bmod 2\pi$$

and

$$\Delta_r w_t^0 = w^0(t+\Delta t_r) - w^0(t) \quad .$$

Then, for $r \geq R_0$

$$\frac{1}{\Delta t_r^{\ell}} \sum_{n\neq 0} \frac{c_n(\widetilde{\Delta}y_t^0(r))}{c_0(\widetilde{\Delta}y_t^0(r))} \leq \frac{2d(\widetilde{\Delta}y_t^0(r))}{\Delta t_r^{\ell}} \tag{125}$$

Using (124), it can easily be shown that for any $\varepsilon > 0$, there exists a positive integer $r(\varepsilon)$ such that

$$\frac{1}{\Delta t_r^{\ell}} \sum_{n\neq 0} \frac{c_n(\widetilde{\Delta}y_t^0(r))}{c_0(\widetilde{\Delta}y_t^0(r))} \leq \varepsilon \qquad \forall\, r \geq r(\varepsilon)$$

Thus

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t^{\ell}} \sum_{n\neq 0} \frac{c_n(\widetilde{\Delta}y_t^0)}{c_0(\widetilde{\Delta}y_t^0)} = 0$$

for any continuous sample function w^0.

To prove (122), we use the assumptions that $p_{x(t)\,\Delta y_t}(\alpha|\beta)$ is bounded for all β and α. Let M be an upper bound. Then rewriting equation (118) for the particular sample path chosen, we have

$$p_{x(t)|\tilde{\Delta}y_t}(\alpha|\tilde{\Delta}y_t^0) = p_{x(t)|\Delta y_t}(\alpha|\tilde{\Delta}y_t^0)$$

$$+ \frac{\sum_{r\neq 0} c_r/c_0}{1+\sum_{r\neq 0} c_r/c_0}\, p_{x(t)|\Delta y_t}(\alpha|\tilde{\Delta}y_t^0)$$

$$+ \sum_{n\neq 0} \frac{c_n/c_0}{1+\sum_{r\neq 0} c_r/c_0}\, p_{x(t)|\Delta\tilde{y}_t}(\alpha|\tilde{\Delta}y_t^0+2n\pi)$$

Thus

$$\left| p_{x(t)|\tilde{\Delta}y_t}(\alpha|\tilde{\Delta}y_t^0) - p_{x(t)|\Delta y_t}(\alpha|\tilde{\Delta}y_t^0) \right| \leq$$

$$2M\left[\frac{\sum_{n\neq 0} c_n/c_0}{1+\sum_{n\neq 0} c_n/c_0}\right] = o(\Delta t^{\ell}) \quad \forall\, \ell \geq 0 \qquad \blacksquare$$

We note that Theorem 9 may still be true even if equation (120) is not satisfied. An examination of the proof shows that all we require is the following : let

$$\mathcal{G}(\alpha^2,\xi) = \int_{-\infty}^{+\infty} e^{-\alpha^2 m^2(u,t)+\xi m(u,t)}\, p_{x(t)}(u)\, du$$

Then we must have

$$\sum_{n\neq 0} \exp - \frac{n\pi(\ n\pi - 2\widetilde{\Delta} y_t^0)}{\ \ t)\ \Delta t}\ \mathscr{G}\left[\frac{\ \Delta t}{2q(t)}, \frac{(2n\pi + \widetilde{\Delta} y_t^0)}{q(t)}\right] = o(\Delta t^{\ell})$$

$$\forall \ell \geq 0 \qquad (126)$$

For instance

$$\mathscr{G}(\alpha^2, \xi) \sim K_1 e^{K_2(\xi+c)^2} + K_3 \xi^r \ ,$$

for given c , K_1 , K_2 , K_3 , which depend only on α and are bounded as $\alpha \to 0$, will satisfy (126) . Thus, for example, Theorem 9 holds if m is linear and $p_{x(t)}$ is normal.